

分类号_____

密级_____

UDC _____

编号_____

中国科学院研究生院

硕士学位论文

空间科学虚拟观测台体系结构研究

Architecture Research of Virtual Space Science Obsertatory

傅衍杰

指导教师_____邹自明 研究员_____

指导教师单位名称_____中国科学院空间科学与应用研究中心_____

申请学位级别_____硕士学位_____ 学科专业名称_____计算机应用_____

论文提交日期_____2011 年 5 月_____ 论文答辩日期_____2011 年 6 月 9 日_____

培养单位_____中国科学院空间科学与应用研究中心_____

学位授予单位_____中国科学院研究生院_____

答辩委员会主席_____

独 创 性 声 明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含任何其他他人已发表或撰写过的材料，也不包含为获得其它教育机构的别种学位或证书而大量使用过的材料。与我一同工作的人对本研究所做的任何贡献已在论文中作了明确的说明并表示谢意。

签名：_____ 日期：_____

关于论文使用授权的说明

本人完全了解培养单位有关保留、使用学位论文的规定，即：培养单位有权保留送交论文的复印件，允许论文被查阅和借阅；培养单位可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。

签名：_____ 见证人：_____ 日期：_____

摘 要

人造卫星、地面台站、探测气球和探空火箭等探测手段的发展,地球探测、月球探测、行星际探测、太阳银河及河外观测等空间计划的推进,使空间科学科学研究进入“数据雪崩”的时代。与此同时,数据网格、云计算等新技术的涌现,使得在协同环境下实现大规模分布式数据的共享成为可能。在这样的背景下,旨在实现数据互操作的虚拟观测台(Virtual Observatory, 简称 VO)应运而生,并迅速受到空间和天文学界的支持和重视。

论文首先简要介绍了虚拟观测的产生背景、概念定义、应用需求及其发展历程,总结了当前虚拟观测在天文学和空间科学两个学科领域的研究进展,同时面向空间科学 VO 体系结构,在借鉴国内外研究成果的基础上,提出了一种 VO 体系结构的分类方法,并对空间科学领域的典型 VO 体系结构进行分析。

论文紧密结合我国空间科学应用研究的特点,开展空间科学 VO 体系结构设计理论的探索,详细分析了体系结构设计中的核心问题,提出了空间科学 VO 的基本设计原则,针对空间科学数据特点,设计了空间科学数据管理模型,然后从概念框架、层次结构、交互协议和服务模型四个方面对空间科学 VO 体系结构进行研究,并进行理论总结。

论文依据空间科学 VO 体系结构理论研究的工作成果,以及在传统数据应用环境建设中遇到的问题,介绍了空间科学 VO 运行引擎的设计动机,然后基于 MVC 设计模式,提出了运行引擎的改进型设计方案。为实现该运行引擎,论文将其划分成框架、控制、视图和模型四个部分,分别进行设计和实现,最终构成一个符合空间科学数据管理模型的上松下紧的漏斗型软件执行结构,并对其工作流的完备性予以推导和证明。

最后,论文基于空间科学 VO 运行引擎,在对空间科学数据资源内容体系和组织架构充分调研的前提下,设计和实现了空间科学 VO 的分布式数据库,以及门户服务、资源注册服务、资源发现服务、资源存储与访问服务、资源收割同步服务、学科应用服务等核心服务,并介绍其具体节点部署和科学应用功能。

实践和应用结果表明,基于论文所提出的体系结构建立的 VO 在解决全局资源的实时同步、异构科学应用的高度集成和面向科学问题的数据发现等问题上,

有一定的普适性。

关键词 空间科学，虚拟观测，体系结构，运行引擎，原型实现

Architecture Research of Virtual Space Science Observatory

YanJie Fu (Computer Application)

Directed by Ziming Zou

Abstract

With the increasing development of space explorations technologies such as satellites, ground stations, weather balloons and sounding rockets and the promotion of national space programs like earth exploration, lunar exploration, planetary exploration, galactic and extragalactic observation, the research of space science is heading into the era of "data avalanche ". At the same time, the advent of data grid, cloud computing and other new technologies makes us much easier to achieve large-scale distributed data sharing in a collaborative environment. On the basis of the background, Virtual observatory emerges as the times require, and catches the attention and emphasis of the community of astronomical and space science.

First, the paper first briefly describes the motivation, concept definition, application demand and development process of Virtual Observatory (VO), summarizes the current research progress of virtual observatory in astronomy and space science, proposes a classification method of Virtual Observatory architecture, and presents case study on the architecture of some typical Virtual Space Science Observatory.

Moreover, based on the characteristics of applied research in space science, the paper analyzes the core problems of architecture design in detail, proposes five basic design principles for Virtual Space Science Observatory(VSSO), designs a data management model for space science, and finally presents an architecture for VSSO from conceptual framework, hierarchy, interaction protocols and service model.

Besides, according to the result of VSSO theoretical research and the problems we encounter in satellite information system, the paper introduces the motivation of VSSO running engine, designs VSSO running engine based on MVC pattern. To implement the running engine, we divide VSSO running engine into four parts:

framework, controller, viewer and model, ultimately form a reusable running engine of software implementation structures specifically for space science data model.

Finally, on the basis of VSSO running engine, we design and implement VSSO systematic prototype including distributed database, portal service, resource registration service, resource discovery service, resource storage and access service, resource harvest and synchronization service, single-sign-on service, security service, science application service, describe the deployment of VSSO and present its application on space weather model computing,

The practice and application results indicate that a VO established on the basis of the VSSO architecture proposed by this paper has certain universality in solving VO resources global discovery, metadata real-time synchronization, resource interoperability and data discovery based on space event association mining.

Keywords: space science, virtual observatory, architecture, running engine, prototype implementation

目 录

空间科学虚拟观测台体系结构研究	I
摘 要	III
Abstract.....	V
目 录	VII
图索引	XI
表索引	XIII
第一章 虚拟观测产生的背景及其发展	1
1.1 虚拟观测产生的背景	1
1.2 虚拟观测的概念定义	4
1.2.1 概述	4
1.2.2 虚拟观测的科学目标	5
1.2.3 虚拟观测的技术目标	6
1.2.4 虚拟观测系统的特征	7
1.3 虚拟观测的国内外发展态势	8
1.4 虚拟观测的应用需求	11
1.5 虚拟观测的相关技术	13
1.5.1 可扩展标记语言	13
1.5.2 网格计算	14
1.5.3 云计算	15
1.5.4 分布式数据库	16
1.5.5 Web Service	17
1.5.6 科学可视化技术	17
第二章 虚拟观测的研究进展	20
2.1 概述	20
2.2 天文虚拟观测的研究现状	20
2.2.1 英国天文网格	22
2.2.2 澳大利亚虚拟天文台	23

2.2.3 欧洲天体物理虚拟天文台	23
2.2.4 日本虚拟天文台	24
2.2.5 美国国家虚拟天文台	25
2.2.6 中国虚拟天文台	26
2.2.7 加拿大虚拟天文台	27
2.3 空间科学虚拟观测的研究现状	28
2.3.1 虚拟太阳观测台(VSO)	28
2.3.2 虚拟空间物理观测台(VSPO)	29
2.3.3 虚拟日地观测台(VHO)	30
2.3.4 虚拟宇宙线观测站(ViCRO)	30
2.3.5 虚拟磁层观测台(VMO)	31
2.3.6 虚拟电离层、热层、中间层观测台(VITMO)	32
2.3.7 虚拟辐射带天文台(ViRBO)	32
2.3.8 虚拟高能粒子观测台(VEPO)	33
第三章 虚拟观测台体系结构的分类研究与典例分析	35
3.1 概述	35
3.2 空间科学虚拟观测台的体系结构分类	35
3.3 VSO 体系结构分析	35
3.4 VMO 体系结构分析	37
3.5 VITMO 体系结构分析	38
3.6 VSPO 体系结构分析	39
3.7 VHO 体系结构分析	40
3.8 IVOA 体系结构分析	40
第四章 空间科学虚拟观测台体系结构的设计理论研究	42
4.1 概述	42
4.2 核心问题分析	42
4.3 基本设计原则	44
4.4 数据模型设计	45
4.5 概念框架	49

4.6 层次结构与协议	50
4.7 服务模型	51
第五章 空间科学虚拟观测台的运行引擎设计与实现	54
5.1 概述	54
5.2 运行引擎的设计动机	54
5.3 运行引擎的基本思想	56
5.3.1 面向计算的 Model 优化方案	57
5.3.2 面向存取的 Model 优化方案	58
5.4 运行引擎的设计与实现	58
5.4.1 框架	59
5.4.2 控制与视图及其完备性证明	60
5.4.3 模型	62
第六章 空间科学虚拟观测台的原型设计和实现	65
6.1 概述	65
6.2 数据模型与数据库设计	65
6.2.1 数据资源内容体系	65
6.2.2 数据资源组织架构	66
6.2.3 数据库设计	68
6.3 核心服务设计	70
6.3.1 门户服务设计	72
6.3.2 资源注册服务设计	73
6.3.2 资源发现服务设计	78
6.3.3 资源存储与访问服务设计	80
6.3.4 元数据收割同步服务设计	82
6.3.5 支持一站式访问的单点登录服务设计	84
6.3.6 监控服务设计	86
6.3.8 安全服务设计	87
6.3.9 学科应用服务设计	87
6.4 系统部署	91

6.5 科学应用	93
6.5.1 空间科学数据服务	93
6.5.2 基于空间事件关联的数据发现	94
6.5.3 科学数据的在线可视化	95
6.5.4 空间科学数据在线阅读与转换	96
6.5.5 空间科学模式计算及其可视化	97
第七章 结语	99
参考文献	101
硕士期间发表文章目录	104
致谢	105

图索引

图 1- 1 空间科学数据应用环境的发展历程	2
图 1- 2 以数据为驱动的科学研究方式	3
图 1- 3 NASA - The Heliophysics Great Observatory.....	3
图 1- 4 空间环境物理因子相互作用框架	6
图 1- 5 SPASE 数据模型体系结构	10
图 1- 6 PDS 数据模型体系结构	11
图 1- 7 太阳高能粒子爆发时太阳风扰动演变过程(1997-5-15)	12
图 3- 1 虚拟太阳观测台体系结构.....	35
图 3- 2 虚拟磁层观测台体系结构.....	37
图 3- 3 虚拟电离层、热层、中间层观测台体系结构.....	38
图 3- 4 虚拟空间物理观测台体系结构.....	39
图 3- 5 虚拟日地观测台体系结构.....	40
图 3- 6 IVOA 虚拟观测台 0 级体系结构	41
图 3- 7 IVOA 虚拟观测台 1 级和 2 级体系结构	41
图 4- 1 空间科学数据公共数据模型	46
图 4- 2 数据元素、专业库和逻辑库的关系	47
图 4- 3 基于公共数据模型的日地空间物理数据资源组织架构	48
图 4- 4 基于公共数据模型的行星科学数据资源组织架构	48
图 4- 5 空间科学虚拟观测台的概念框架	49
图 4- 6 空间科学虚拟观测台节点与要素的耦合	49
图 4- 7 空间科学虚拟观测台的层次结构	50
图 4- 8 资源层、服务层和应用层之间的交互	51
图 4- 9 服务透明切换策略	53
图 4- 10 空间科学虚拟观测台服务模型	53
图 5- 1 空间科学 V0 运行引擎设计方案	59
图 5- 2 运行引擎的控制与视图运行时序	61
图 5- 3 空间科学 V0 运行引擎的 Model 结构图	63
图 6- 1 数据资源结构图	67
图 6- 2 数据资源存储方式	68
图 6- 3 空间科学虚拟观测台数据库逻辑设计	69
图 6- 4 空间科学虚拟观测台数据库 ER 图	70
图 6- 5 资源注册阶段的数据描述信息状态转换图	74
图 6- 6 数据资源注册服务流程设计	75
图 6- 7 在线数据的资源录入	76
图 6- 8 离线数据的资源录入	76

图 6- 9 数据资源审核	77
图 6- 10 数据资源发布	78
图 6- 11 数据资源发现的数据流图	78
图 6- 12 资源发现服务设计	79
图 6- 13 资源存储与访问服务设计	81
图 6- 14 资源访问服务与监控服务的耦合	81
图 6- 15 推拉式的数据收割同步策略	83
图 6- 16 数据收割服务设计	84
图 6- 17 基于票据的单点登录服务设计	85
图 6- 18 单点登录流程示意图	85
图 6- 19 监控服务设计	86
图 6- 20 基于空间天气事件的资源关联发现	88
图 6- 21 空间科学事件关联发现三要素	89
图 6- 22 MISIE 大气模式计算及科学可视化服务	90
图 6- 23 空间科学虚拟观测台节点分布	92
图 6- 24 空间科学虚拟观测台拓扑结构	92
图 6- 25 基于核心属性、语义与分类导航的数据发现	93
图 6- 26 数据检索与定位	94
图 6- 27 空间天气质子事件检索	94
图 6- 28 质子事件记录可视化分析结果	94
图 6- 29 质子事件关联数据检索	95
图 6- 30 嫦娥月球微波数据和怀柔太阳观测数据的在线可视化	95
图 6- 31 CDF 在线阅读和多格式互转	96
图 6- 32 SPASE 元数据标签阅读器	97
图 6- 33 美国海军实验室 MISI 大气模型在线计算及三维可视化	97
图 6- 34 IGRF 地磁模型在线计算及其三维可视化	98

表索引

表 1 虚拟观测项目一览表.....	22
表 2 ViRB0 参建单位	33
表 3 空间科学数据核心属性解释表	46
表 4 空间科学 V0 数据库设计模型与元数据划分	68
表 5 空间科学虚拟观测台核心服务分析.....	71

第一章 虚拟观测产生的背景及其发展

1.1 虚拟观测产生的背景

自古以来，了解太空，探索地球以外的物质，一直是人类不懈追求的梦想。20 世纪以来，短波无线电远程通信试验成功，电离层的发现，宇宙线的观测，磁暴和电离层暴 27 天重现性与太阳自转有关的发现，以及等离子体振荡的发现等促进了理论研究的发展，为空间科学技术的形成奠定了基础^[1]。50 年代以后，在大量地面台站、气球和火箭观测及长期理论研究的基础上，迫切要求各相关学科之间密切配合，要求全球性的协同观测以及发展新的探测手段。1957 年，苏联首次发射了人造地球卫星，这标志着人类进入了空间时代。随后，人类对近地空间环境进行了大量的普查，发现了地球辐射带、环电流，证实了太阳风、磁层的存在，发现了行星际磁场的扇形结构和冕洞等；月球探测器和“阿波罗”飞船载人登月，对月球进行了探测和综合性研究；行星际探测器系列对行星进行了探测，并由对内行星发展到内外行星的探测；天文观测卫星系列对太阳、银河辐射源、河外源。

在取得上述进展的同时，空间生命科学也相应地迅速发展起来。例如研究人在空间长期生存的一系列问题，包括在失重、超重、高能辐射、节律改变等条件下人体的适应能力等；空间生物学、医学和生保系统的研究也取得了很大的进展；关于地外生命也在进一步探索。至此，空间交叉领域的蓬勃发展催生了一个与现代化观测技术结合的新学科：空间科学技术^[2]。

空间科学技术^[3]是指利用航天器研究发生在日地空间、行星际空间及至整个宇宙空间的物理、天文、化学及生命等自然现象及其规律的科学。空间科学技术以航天技术为基础，包括空间飞行、空间探测和空间开发等几个方面。空间科学技术的出现，开辟了人类探索外层空间活动的新时代。经过近半个世纪的迅速发展，人类航天活动取得了巨大成就，极大地促进了生产力的发展和社会的进步，产生了重大而深远的影响。空间科学技术已成为当今世界高技术群中对现代社会最具影响的高技术之一，是世界各国现代化建设的重要内容。

从 20 世纪 90 年代开始，空间科学技术经历了革命性的变化。这一变化是由

前所未有的信息技术革命所推动的。卫星设计和制造工艺的发展，大型望远镜和传感器探测精度的提高，大规模高性能计算和模拟技术的出现，遥感通讯、信息系统和网络技术的进步，使得空间探测一跃成为了空间科学研究的基础，并迅速被世界各国视为未来太空战略的制高点和试验场。

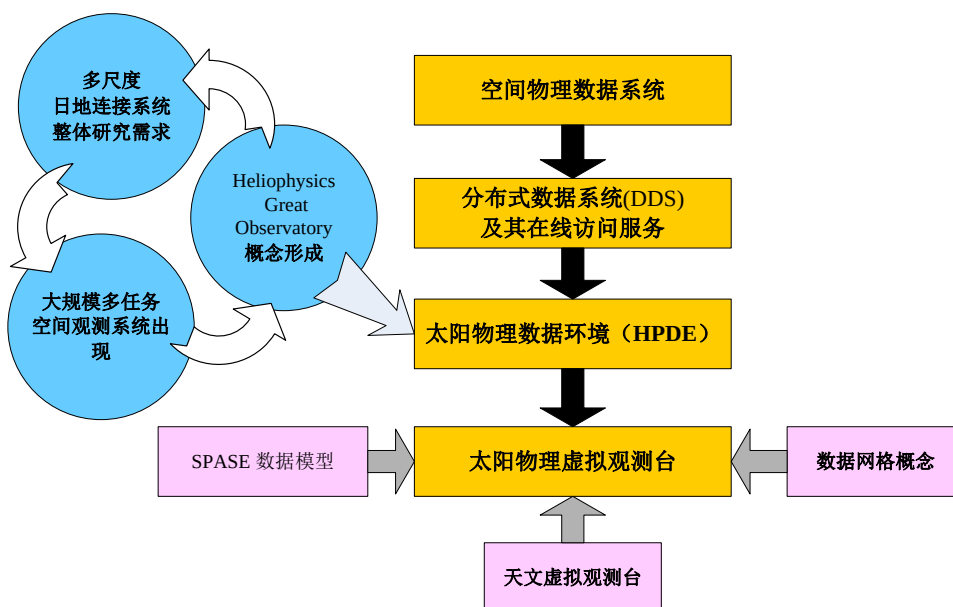


图 1-1 空间科学数据应用环境的发展历程

如同在计算机工业中，计算能力随时间指数增长的摩尔定律一样，在过去几十年的技术发展，使得空间天文探测的数据收集能力也遵循着摩尔定律。国内外的科学卫星的发射^[4]，如嫦娥系列、风云系列、实践系列、SOHO 系列、探索者系列等，以及地面台站的建设，如子午工程、LAMOST 计划、国际日地空间物理大观测台、国际 SuperDARN 地球两极电离层雷达观测计划、日不落联合观测计划等，使得科学研究进入“数据雪崩”的时代。与此同时，空间科学数据应用环境也经历了从单机式环境，到分布式数据应用环境，再到虚拟观测的变更（见图 1-1）。

在空间科学方面，现有的以及即将在十二五“空间科学先导专项”启动的空间科学探测项目，每天都将会产生 GB 甚至 TB 级别的数据。其中，空间科学先导专项，发射五个系列卫星，重点针对黑洞的性质及极端条件下物理规律、暗物质的性质、空间环境下的物质运动规律和生命活动规律、太阳爆发等太阳活动对地球空间环境的影响和检验量子力学完备性等方面开展研究，并把科学卫星上天

后的科学数据应用，作为构成空间科学任务从孵育、前期准备、技术攻关到工程研制、成果产出的完整链条的核心部分之一。

在天文观测方面，哈勃空间望远镜（HST）每天大约产生 5GB 的数据；我国正在建造的大天区面积多目标光纤光谱望远镜（LAMOST）也将每天产生 3GB 的数据；而美国计划建造的“大口径综合巡天望远镜（LSST）”，又称为“暗物质望远镜”，每天的观测数据将达到 18TB 的量级^[5]。

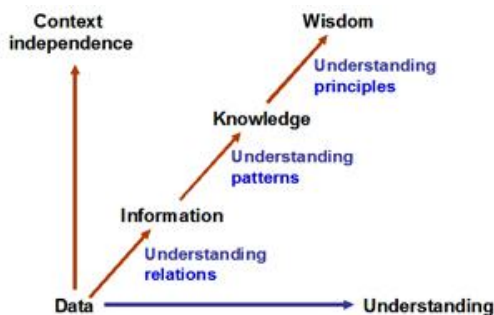


图 1-2 以数据为驱动的科学研究方式

巨大的海量的观测数据产出，使得在空间探测历史上，人类进入一个数据爆炸的时代。海量科学探测数据还使得传统的“猜测、推理、归纳、观测、验证”的研究方式，变为数据驱动的现代化研究方式（见图 1-2）——“观测、计算、挖掘、分析、验证”。与此同时，高性能计算、数据网格、预报反演、模拟仿真、科学可视化等新技术的涌现，使得在协同环境下，实现大规模分布式数据的共享和应用成为可能。

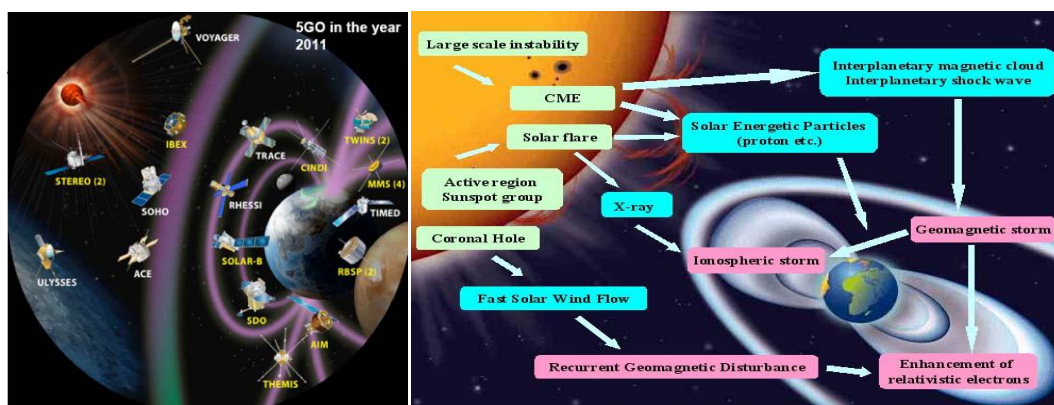


图 1-3 NASA - The Heliophysics Great Observatory

在这样的背景下，虚拟观测的概念应运而生（见图 1-3），并迅速受到空间

和天文学界的重视。虚拟观测思想的本质是，缩短空间探测的空天段和地面段的“科学距离”，推进地面测控、空间探测、数据应用、先进技术的深度一体化，实现科学数据应用价值的最大化。

1.2 虚拟观测的概念定义

1.2.1 概述

什么是虚拟观测？我们可以先用一个形象的比喻，来阐述这个问题。如果我们首先把在地面段中全球各地的地面站和测控船的远程操作，虚拟化到一个集中式的运控平台，就等于为工程师建立了虚拟测控中心；再对低轨道、中轨道和同步卫星轨道上的各种通讯卫星、遥感卫星、科学卫星、气象卫星等卫星载荷的观测数据，按照日地空间三维网格坐标结构、空间观测物理要素分类和观测时间坐标的五维体系，进行清洗、归档和布局，便可以构建一个虚拟的数字空间环境；而根据科学工作者的要求，发现和挖掘特定空间观测区域的科学数据，就仿佛是在使用一个虚拟空间站或卫星上的虚拟观测设备；如果再根据科学应用的要求，设计和研发空间模式计算、空间天气预报、空间环境模拟、空间碎片仿真等工具，这就相当于拥有了虚拟的科研实验室。在这种情况下，虚拟运控中心、虚拟空间环境、虚拟探测设备和虚拟实验室所组成的，便是一个真正意义上的虚拟观测台（Virtual Observatory）。

然而，到目前为止，尚未有一个权威的组织全面地给出了虚拟观测的定义。当前，科学界对虚拟观测定义或是从软件的视角，或是科学应用内涵的视角去看待 VO。2004 年 10 月，美国宇航局（NASA）曾以白皮书方式提出 VO 的概念——“*A Virtual Observatory (VO) is a suite of software applications on a set of computers that allows users to uniformly find, access, and use resources (data, software, document, and image products and services using these) from a collection of distributed product repositories and service providers. A VO is a service that unites services and/or multiple repositories.*”^[6]从科学价值和科学应用的角度出发，中国科学院空间科学与应用研究中心邹自明研究员也给出了虚拟观测的定义^[7]：“虚拟观测是依赖于人类掌握的最新空间科学知识，面向天、地基观测产生的海量空间科学数据资源，采用先进的信息技术，以解决海量空间科学数据协同处理、

分布式集成与互操作、智能化科学应用、日地空间系统可视化虚拟表达等核心问题为目标的日地空间科学数据应用环境。”

1.2.2 虚拟观测的科学目标

目前，空间科学和天文学界共同确定虚拟观测的主要科学目标是：

(1) 日地空间和行星系统的空间分区与虚拟表达

将科学卫星和航天器的多物理要素的载荷探测数据，以及地面台站的多波段的巡天数据，通过虚拟观测技术，基于数据时空模型，表达到精确的空间网格坐标系中，以此来实现人类在不同波段、不同辐射环境上重构出多层次的大尺度的系统的外太空空间环境。

(2) 新型空间事件和稀有天体的发现

通过卫星、航天器或者地面台站的巡天观测，来发现新型空间事件或稀有天体（如高红移类星体、褐矮星等）是目前空间天文探测的主要科学问题之一^[5]。以天文探测为例，如果从统计学上审视天文观测，假如某种稀有天体或现象在为人观测所捕获概率是百万分之一或一亿分之一，那么就需要几百万或几亿个观测数据样本才有可能发现。这说明，要进行宇宙探索，寻找稀有的未知类型天体，必须面对海量观测数据。而虚拟观测技术将会利用其独有的数据资源和计算资源促进新的空间天文的科学发现。

(3) 空间环境描述、缓变和扰动规律研究

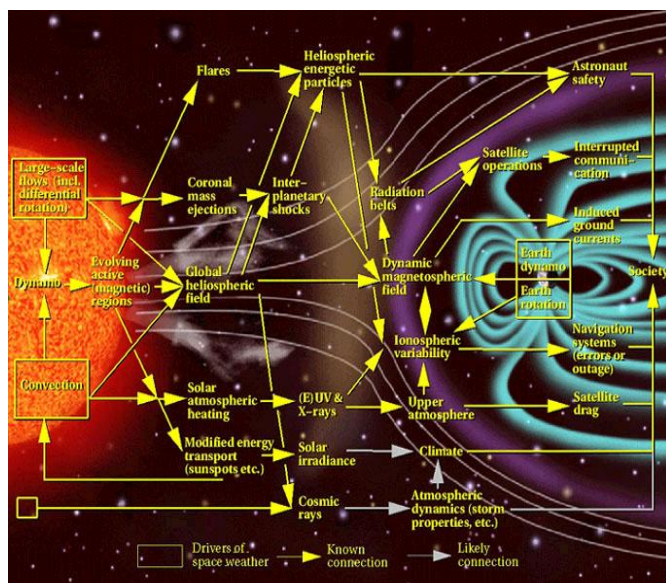


图 1-4 空间环境物理因子相互作用框架

空间在轨卫星和航天器的设计能力是一个国家空间战略执行力的表现。其中，空间环境研究是在轨管理和航天器设计的重要保证。因而，研究空间环境的定量描述方式、缓变规律和异常扰动机制（见图 1-4），对于开展空间事件爆发预警和空间导航具有重大意义。

(4) 促进新科学领域的兴起

虚拟观测的出现使得科学研究方式从传统的理论驱动的“猜想、推理、归纳、观测、证明”转变为现代化的数据驱动的“观测、计算、挖掘、分析、验证”的新研究模式。虚拟观测还将会促进空天和地面多载荷、多波段空间天文联合探测的发展，从深层次来探索宇宙。虚拟观测还将促进统计空间科学和统计天文学的兴起，如空间辐射环境的静态统计模型、长期缓变经验模型、扰动爆发统计模型、宇宙大尺度结构和银河系结构的定量分析、天体和空间环境观测样本的建立与研究等^[5]。

(5) 基于空间天文海量观测数据的知识发现

统计空间科学和统计天文学的出现，加快了数据挖掘、知识发现、人工智能、统计学和概率论等计算机和数学理论分支的引入。如何在一个海量分布式的数据应用环境下，利用高性能计算、物理理论和数学算法，实现空间天文知识的发现成为了一个当前非常热门的交叉应用新研究领域。

1.2.3 虚拟观测的技术目标

目前，空间科学和天文学界共同确定虚拟观测的主要技术目标^[8]是：

- (1) 统一科学数据模型
- (2) 资源的无缝链接和互操作
- (3) 分布式资源和节点的快速融合、共享和访问
- (4) 资源全局、高效、多维度的发现
- (5) 资源在分布式环境下的同步
- (6) 资源与学科应用高度集成和整合
- (7) 研究科学工作流，实现科学数据协同处理
- (8) 面向数据溯源和数据组装的数据管道技术

(9) 科学计算结果的可视化技术

(10) 系统的逻辑统一性和透明性

1.2.4 虚拟观测系统的特征

(1) 异构性

虚拟观测系统属于分布式系统，其异构性体现在 6 个方面，分别是：①数据格式异构；②数据管理模型异构；③数据访问协议异构；④服务访问接口异构；⑤节点运行环境异构；⑥安全协议异构。

(2) 物理分布性

虚拟观测系统一般是由多个节点，按照特定的拓扑结构连接交互而成的分布式系统。这些节点分布在不同的物理位置，不同的节点都可能拥有不同的资源（数据、计算、存储、应用等）。

(3) 可扩展性

虚拟观测的根本目的是为了最大限度的实现数据资源的整合和共享，因而虚拟观测系统不应该仅仅定位在少量数据资源的局域系统，而应该把目标定位在整合跨城域，甚至跨国际的广域性的数据资源。通过快速的整合异构异地的数据资源，使之融合并成为数据资源提供者的一部分。通过可扩展性，VO 可以让无数的数据资源拥有者变成数据资源提供者。这些数据资源提供者可以分为两类：①利用虚拟观测系统中的已有节点对外提供数据资源服务，即资源提供者不提供存储力，仅提供数据资源；②资源提供者作为新的数据节点，对外提供资源访问等服务。前者改变了数据资源在拓扑的分布；后者不但改变数据资源的分布，还改变系统的网络拓扑。

(4) 透明性

虚拟观测要做的是把物理结构上分布式的资源变成逻辑上的统一透明^[9]。虚拟观测的透明性包含数据访问透明性、服务访问透明性和应用访问透明性。因而，物理分布、逻辑整体和站点自治是不可脱离，相辅相成的。系统的透明性可以分为 4 个方面：

- 命名的透明性：网格需要维护异构环境中的元数据和物理数据实体的信息，为用户提供统一的逻辑视图。而事实上，同一个集合或者同一个类

别的数据可能分布在异构的存储系统中。为实现命名的透明性，网格需要管理不同的名字和属性，甚至它们相互之间的关系，并把高层命名映射为异构分布式存储环境中的底层命名和存储操作。

- 定位的透明性：由于网格的物理分布性，这就决定了资源定位的过程是一个分布式的查找过程。而网格的目的在于让用户体会到，一打开水龙头就能使用来自不同地方的水资源。
- 协议的透明性：网格应该支持异构数据资源的访问，网格为用户提供统一的访问接口，为实现用户提出的访问请求，网格能够自动判断底层异构的访问协议，并自动进行转换，而在高层访问协议将会是一个统一的描述。
- 时间的透明性：受带宽和异构的影响，分布式访问和寻址，将会耗费大量的时间，时间的消耗必然导致延迟，如何通过 workflow 安排和本质效率提高两个方面去减少用户体会。

(5) 健壮性

虚拟观测系统应该具有健壮性，也叫自愈能力^[10]。虚拟观测系统的这种拓扑自组织能力，能够针对不断变动的网络拓扑和数据资源分布，进行同步和报告，以支持资源发现、监控、安全、备份等服务。同时，越来越多的数据资源和资源提供者的加入，大大加大虚拟观测系统管理运行的难度和复杂性。由于，在日常生活中，这些资源对外提供的服务发生问题的概率很高。因而，虚拟观测系统需要能够对用户屏蔽系统故障，动态的适应、报告和纠正系统错误。不仅如此，系统还可以通过监控服务，记录和分析系统状态，了解服务和应用的健康情况。当任一节点的服务出现问题时，系统能自动将该服务的访问负载，切换到附近节点的同类服务上，以提高服务的健壮。

1.3 虚拟观测的国内外发展态势

大规模多层次的科学观测为人类的科学研究提供了丰富的数据资源，由数据驱动的知识发现已成为人类开展科研活动的主要方式之一。近二十年来，基于互联网络技术发展建立的科学数据共享服务系统为科学数据的发现、访问、下载提供了极大方便，大大促进了科学数据的应用。然而，数据驱动的科研活动正处于

由单学科研究向跨学科综合研究方式的转变过程中,传统的以独立方式对外提供数据访问与下载服务为主要特点的科学数据共享服务体系架构已经完全不能满足这种跨学科综合研究对数据大规模综合应用与访问的需求,发展以满足跨学科数据综合应用访问需求为核心目标的全新数据共享服务体系正在成为国际科学数据管理界和信息技术界的重大课题之一。

近年来,建立以科学虚拟观测台(VO)为主要特点的科学的科学数据共享服务体系(学科数据环境),满足科学研究对数据综合应用与访问的需求,已基本得到国际科学界的认可。虚拟观测台由分布式数据管理系统发展而来,数据网格(DataGrid)技术为其主要支撑技术。虚拟观测台概念首先出现在天文学领域,2003年成立的国际虚拟天文台联盟(IOVA)第一次在国际上提出建立国际虚拟天文观测台,旨在促使分布在世界各地的天文数据档案进行合作,使其能够以集成地方式向用户提供综合的天文数据档案综合应用访问与互操作。目前,全球已有十六个国家参加了国际虚拟天文台联盟(IOVA)的活动。受此影响,地球科学、空间科学界也纷纷提出建立各自的虚拟观测台计划,启动了一大批与此相关的科学项目。2007年7月在意大利佩鲁贾举行的第24届国际大地测量学与地球物理学联合会(IUGG)大会正式启动的国际电子地球物理年(eGY)活动把推动发展地球科学领域的科学虚拟观测台作为其最主要的两个任务目标之一。

从上世纪九十年代开始,随着日地空间物理探测数据的海量增长,日地空间物理学界就提出了建立日地空间物理数据环境的动议,美国航空航天管理局(NASA)是事实上的主要的推动机构,为了推动日地空间物理数据环境体系(HPDE)的发展,早在2002年1月,NASA“与日同在”计划(Living With Star)就成立了专门的数据环境工作组。为了解决不同数据系统之间的互操作的技术标准问题,NASA联合欧洲的相关空间科学研究机构于2003年10月建立了国际空间物理档案研究与抽取联盟(SPASE),该联盟目前已发布了三个SPASE标准,当前的版本为Version 1.2.0。2004年10月,NASA“与日同在”计划组织召开的国际会议,就如何建立日地空间物理虚拟观测台达成一致,形成了“空间与太阳物理虚拟观测台框架”报告,该框架极大地推动了国际日地空间物理虚拟观测台的发展。与此同时,NASA按照先分学科领域后集成的技术路线,启动了日地空间物理多个分学科领域的科学虚拟观测台项目。到目前为止,正式启动的虚拟观

测台项目包括：全球极光图像访问系统（GAIA）、虚拟太阳观测台（VSO）、虚拟空间物理观测台（VSPO）、虚拟日地观测台（VSTO）、欧洲太阳观测台网格（EGSO）、虚拟日球观测台（VHO）、虚拟磁层观测台（VMO/G）、虚拟磁层观测台（VMO/U）、虚拟辐射带观测台（ViRBO）和虚拟电离层热层中性层观测台（VITMO）。其中的绝大部分项目申明采用 SPASE 发布的数据管理模型与元数据标准。

VO 的科学应用价值很大程度上反映在模式应用上。目前，在由美国国家空间科学数据中心、美国国家地球物理数据中心、美国太阳数据分析中心、美国一致模型中心（CCMC）、美国西南研究院(SwRI)等研究机构组成的空间物理数据与模式联盟（Heliophysics Data Model Community）正致力于发展研制电离层、大气、等离子体、地磁场、太阳和行星际空间等模型，通过事件分析、实计算检测和评估模式的可用性，为科学用户提供应用支持，同时开展空间模式框架和融合的研究。科学家通过模式研究，深入了解各圈层的完整图像和物理过程、各圈层之间的耦合机制和传输过程，需要科学数据模型来支持科学问题的发现。因而，在数据模型方面，空间物理归档检索抽取标准[11]（SPASE，见图 1- 5）、行星科学系统[12]（PDS，见图 1- 6）和 IVOA VOTable[13]等学科组织以 Workshop 的方式不断推进数据标准的统一化，以实现虚拟观测中高效快速的数据互操作。

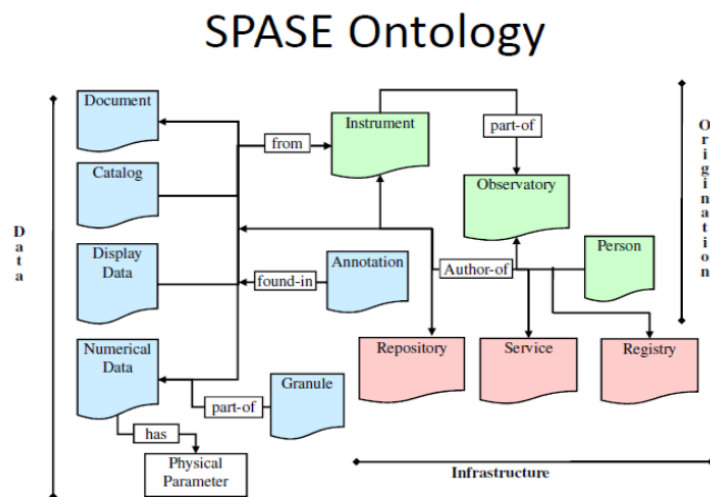


图 1- 5 SPASE 数据模型体系结构

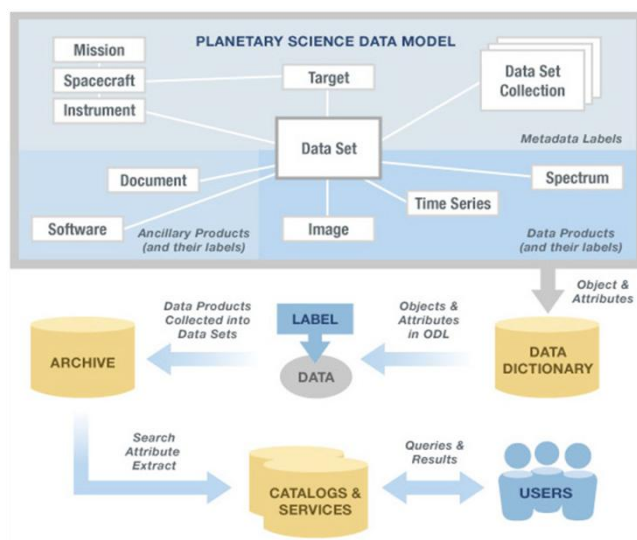


图 1- 6 PDS 数据模型体系结构

与此同时，在天文学领域，国际虚拟天文台联盟（IVOA）已经拥有 19 个成员国（截至 2011-4-1 统计数据）。IVOA 目前正在着重在学科应用、数据访问标准和协议、数据管理模型、资源注册、安全认证、特定语义检索等方面建设天文学的虚拟观测系统。

1.4 虚拟观测的应用需求

随着大规模多层次的科学观测为人类的科学研究提供了丰富的数据资源，由数据驱动的知识发现已成为人类开展科研活动的主要方式之一。近二十年来，基于互联网络技术发展建立的科学数据共享服务系统为科学数据的发现、访问、下载提供了极大方便，大大促进了科学数据的应用。

然而，数据驱动的科研活动正处于由单学科研究向跨学科综合研究方式的转变过程中，传统的以独立方式对外提供数据访问与下载服务为主要特点的科学数据共享服务体系架构已经完全不能满足这种跨学科综合研究对数据大规模综合应用与访问的需求，如何解决分布式数据资源集成共享问题和学科应用集成问题，发展以满足跨学科数据综合应用访问需求为核心目标的全新数据共享服务体系正在成为空间科学数据管理与服务所面临的重大课题。

空间科学虚拟观测台（Virtual Space Science Observatory, 简称 VSSO）正是利用互联网实现空间科学数据资源、计算资源、存储资源和知识资源的无缝链接与全面共享，为空间科学研究提供易用、高效、智能的虚拟集成应用环境的最优

平台。它的应用需求源于如下两点：

(1) 分布式数据资源集成、发现与访问的需要

空间科学研究涉及太阳物理和空间物理等多个学科分支，这包括场物理、等离子体、能量粒子、宇宙尘埃、中高层大气等，研究范围从太阳内部到太阳大气，直至行星表面，对某个问题的研究经常需要应用到相关的多个数据集，甚至还需要用到模式输出结果和数值仿真结果。由于不同内容的空间天气数据集分布在不同的地方，数据存储格式各异，采用不同的平台环境，数据访问接口相互独立，数据用户必须花费大量的时间去发现数据、理解数据格式，才能做到对数据的正确使用，浪费了大量的人力和物力。这种以单站单库独立对外提供数据应用服务模式构建的空间科学数据共享服务体系，无法高效率地满足空间天气研究跨数据集的数据应用需求。因此，如何利用互联网环境集成分散在各个研究单位或个人手中的空间科学数据资源，为科学研究者解决全局的数据资源发现与访问，是数据资源集成需要重点解决的问题。

(2) 在线分析服务、模式运算、空间天气预测等学科应用的集成环境的需求

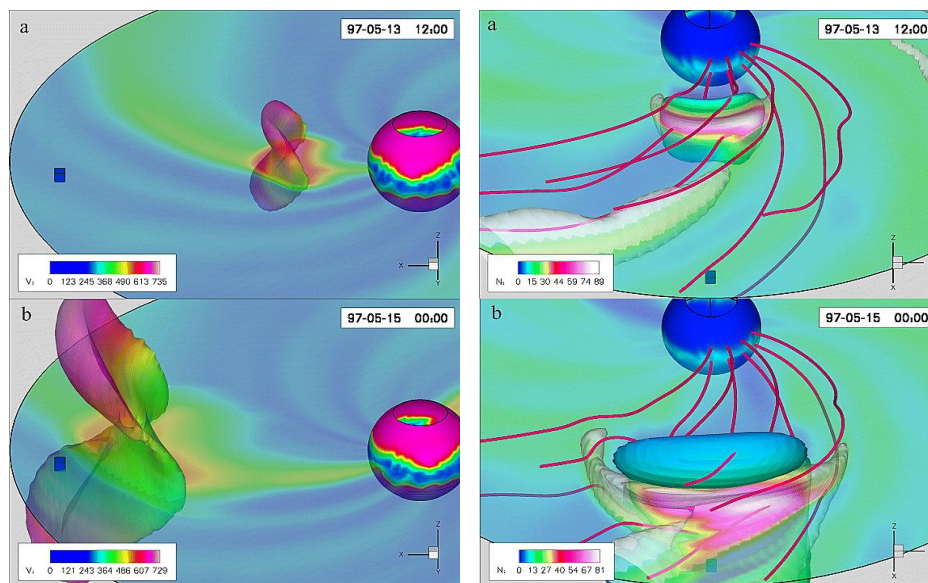


图 1-7 太阳高能粒子爆发时太阳风扰动演变过程(1997-5-15)

当前，空间科学研究问题空前复杂化，往往需要涵盖更大的范围，需要便捷的掌握有关科学事实和数据，在仿真、模式运行与大规模计算成为分析、发现和预测空间科学问题的主要手段时，对某个问题的研究经常需要应用到相关的多类数据，甚至还需要用到模式输出结果和数值仿真结果，由于数据应用工具相对缺

乏，而且通用性差，数据用户在获取数据后使用数据时必须花费大量的时间去查询相关的数据工具，或是基于模式运行需要将数据进行格式转换等二次处理，无法把主要精力投入到知识发现上来，这种仅提供数据服务的空间科学数据共享服务模式，降低了数据的易用性，从一定角度制约了空间天气研究活动的开展，非常需要提出新的数据共享服务体系架构，支持数据用户实现基于物理要素、物理事件、飞行器、时空范围、关键词等要素的综合的数据发现、透明访问和实现数据与在线服务资源无缝链接的全新空间天气数据应用环境系统。

1.5 虚拟观测的相关技术

1.5.1 可扩展标记语言

可扩展置标语言^[14]（eXtensible Markup Language，简称 XML）是一种置标语言，是从标准通用置标语言（SGML）中简化修改出来的。它主要用到的技术包括可扩展置标语言、可扩展样式语言（XSL）、XBRL 和 XPath 等。XML 被广泛用来作为跨平台之间交互数据的形式，主要针对数据的内容，通过不同的格式化描述手段（XSLT，CSS 等）可以完成最终的形式表达（生成对应的 HTML，PDF 或者其他的文件格式）。XML 与 Access、Oracle 和 SQL Server 等数据库不同，数据库提供了更强有力的数据存储和分析能力，例如：数据索引、排序、查找、相关一致性等；而 XML 仅仅是展示数据。XML 与 HTML 的设计区别是：XML 是用来存储数据的，重在数据本身。而 HTML 是用来定义数据的，重在数据的显示模式。

XML 的简单使其易于在任何应用程序中读写数据，这使 XML 很快成为数据交换的唯一公共语言，这意味着程序可以更容易的与 Windows、Mac OS、Linux 以及其他平台下产生的信息结合，然后可以很容易加载 XML 数据到程序中并分析他，并以 XML 格式输出结果。XML 具有六个方面的优势：①XML 可以从 HTML 中分离数据；XML 可以用于交换数据；②XML 可以用于共享数据；③XML 可以用于存储数据；④XML 可以充分利用数据；⑤XML 可以用于创建新的语言。

1.5.2 网络计算

网络是一种元计算技术,它将物理上分布、系统异构的多种计算资源通过高速网络连接起来,共同解决大型应用问题。网络体系结构主要是研究网络系统的基本功能结构及各功能实体间的接口关系。网络不仅实现对各种计算资源的访问,而且实现对所有数据资源的统一访问,还建立了一种有效的数据管理体系结构,将网络计算的数值计算能力和数据网格的数据处理和发布能力有机结合起来,为科学应用在网格中实施资源发现和知识挖掘,并进行高效访问提供支持^[15]。目前网络体系结构^[16]可以分为如下几类:

(1) 面向科学计算及仿真应用的网格

面向科学计算及仿真应用网格^[17]的基本功能,就是利用网格技术将多种复杂的分布式的计算资源组织起来,形成一个分布式、松耦合的巨型计算系统。该计算系统的主要任务是完成超大规模的、复杂的仿真和计算任务,如中长期天气预报、航空航天跨领域的安全仿真等等。

科学计算网格的关键问题是组织、访问和管理计算任务,这是因为科学计算问题一般都是计算密集型问题,其 CPU 的开销是巨大的,因此,如何将一个大问题分解为并发的任务,并将这些任务分配到多个异构的计算系统中去,同时将这些并发的任务有机的组织起来,以尽量小的管理开销达到完成一个共同计算任务的目的,则成为科学计算网格的中心任务。

通常,科学计算网格的体系结构可分为应用层、网格通用服务及本地资源层。科学计算网格至少需要具备三种基本功能:①任务管理,用户通过该功能向网格提交任务、为任务指定所需资源、删除任务并监测任务的运行状态;②任务调度,用户提交的任务由该功能按照任务的类型、所需资源、可用资源等情况安排运行日程和策略;③资源管理,确定并监测网格资源状况,收集任务运行时的资源占用数据。

(2) 数据网格

数据网格^[18]是以大规模数据的共享、存贮、传输及分析为基本特征。初期的数据网格项目就是为了管理高能物理试验的海量数据而建立的,数据网格系统一般都有海量数据的联机采集系统,如高能物理仪器,它可以以 Pbytes/Sec 量级产生数据;高速数据传输,根据不同的应用,数据传输速率从 100M—2.5G 不等;

从地域上讲，可跨世界多个国家和地区。海量数据的共享、访问、控制、管理及传输是数据网格的核心问题。著名的数据网格有 EU Data Grid、GriphyN、Particle Physics Data Grid、TeraGrid、European Data Grid、Japanese Grid Projects 等。

数据网格体系结构中最明显的特征是在网格通用服务层中增加或强化了一些数据库或数据管理服务，如：数据管理（Data Management）、元数据管理（Metadata Management）、对象—文件的映射（Object-File Mapping）、复制管理（Replication Management）、SQL 数据库服务（SQL Database Server）、复制目录（Replica Catalog）等。

数据网格以海量数据的存贮、传输、共享及分析为基本任务，因此数据网格必须如下能力：分析任务的能力，主要完成海量数据的分析，并对分析结果进行可视化处理，以使用户使用；随时掌握网络中资源的能力；执行任务程序的能力，可根据用户需要执行分析程序；任意数据传输的能力，数据传输是数据网格基础通信服务，它对通信能力的要求相当高，突发强；判定和保障服务质量的能力；从错误中恢复的能力等。

(3) 知识网格

知识网格与数据网格的主要区别在于知识网格的数据结构及数据之间的关系更加复杂，它主要反映在如下几个方面：数据对象的来源多样化，数据对象的信息，即属性更加复杂，数据对象的知识，即数据对象之间的关系结构复杂化，难以用传统数据库的二维关系来表达。

目前知识网格属于智能信息处理的范畴，与它相关的内容有语义网（Semantic Web）、知识管理（Knowledge Management）、知识本体（Ontology）、智能主体（Agents）、信息网格、一体化智能信息平台等。

1.5.3 云计算

云计算^[19]（Cloud Computing），是一种新兴的共享基础架构的方法，可以将巨大的系统池连接在一起以提供各种 IT 服务。狭义云计算是指 IT 基础设施的交付和使用模式，指通过网络以按需、易扩展的方式获得所需的资源（硬件、平台、软件）。提供资源的网络被称为“云”。“云”中的资源在使用者看来是可以无限扩展的，并且可以随时获取，按需使用，随时扩展，按使用付费。广义云计算

是指服务的交付和使用模式,指通过网络以按需、易扩展的方式获得所需的服务。这种服务可以是 IT 和软件、互联网相关的,也可以使任意其他的服务。云计算具有如下几个特点:超大规模、虚拟化、高可靠性、通用性、高可扩展性、按需服务和极其廉价。

云计算的服务层次可分为基础设施层、平台服务层以及软件服务层。

(1) 软件服务层—SaaS(Software as a Service, 软件即服务)

SaaS 是一种软件布局模型,其应用专为网络交付而设计,便于用户通过 Internet 托管、部署及接入,即软件提供者将应用软件统一部署在服务器上,客户可以根据实际需求,通过互联网向软件提供者订购所需的应用软件服务。SaaS 供应商不需要再在服务器和带宽等基础设施上浪费自己的资源,专注于具体的软件开发和应用。

(2) 平台服务层—PaaS(Platform as a Service, 平台即服务)

PaaS 是在云基础设施之上提供系统运行的软件平台,例如开发平台、商业部署和应用平台等。PaaS 获取硬件资源的方式对于用户来说是透明。平台服务提供商提供硬件、软件、操作系统、软件升级、安全以及其他应用程序托管等服务内容。目前,大多数提供商限于某种语言和集成开发环境(IDE),如:谷歌的 App Engine 支持 python 以及相应的 IDE。云计算的平台服务对用户屏蔽了操作系统、硬件以及存储的复杂性。

(3) 基础设施服务层—IaaS(Infrastructure as a Service, 基础设施即服务)

基础设施提供商拥有大量的计算资源,例如存储和计算能力,利用虚拟化技术实现了分割、动态调整资源的功能,能够为用户或者服务提供商提供指定规模的系统。

1.5.4 分布式数据库

分布式数据库系统^[20]通常使用较小的计算机系统,每台计算机可单独放在一个地方,每台计算机中都有 DBMS 的一份完整拷贝副本,并具有自己局部的数据库,位于不同地点的许多计算机通过网络互相连接,共同组成一个完整的、全局的大型数据库。分布式数据库系统由分布于多个计算机结点上的若干个数据库系统组成,它提供有效的存取手段来操纵这些结点上的子数据库。分布式数据

库在使用上可视为一个完整的数据库,而实际上它是分布在地理分散的各个结点上。当然,分布在各个结点上的子数据库在逻辑上是相关的。

这种组织数据库的方法克服了物理中心数据库组织的弱点。首先,降低了数据传送代价,因为大多数的对数据库的访问操作都是针对局部数据库的,而不是对其他位置的数据库访问;其次,系统的可靠性提高了很多,因为当网络出现故障时,仍然允许对局部数据库的操作,而且一个位置的故障不影响其他位置的处理工作,只有当访问出现故障位置的数据时,在某种程度上才受影响;第三,便于系统的扩充,增加一个新的局部数据库,或在某个位置扩充一台适当的小型计算机,都很容易实现。然而有些功能要付出更高的代价。例如,为了调配在几个位置上的活动,事务管理的性能比在中心数据库时花费更高,而且甚至抵消许多其他的优点。

1.5.5 Web Service

Web Service^[21]是一种新的 web 应用程序分支,他们是自包含、自描述、模块化的应用,可以发布、定位、通过 web 调用。Web Service 可以执行从简单的请求到复杂商务处理的任何功能。一旦部署以后,其他 Web Service 应用程序可以发现并调用它部署的服务。

Web Service 架构的基本思想,就是尽量把非核心功能交给其他人去做,自身全力开发核心功能。Web Service 平台需要一套协议来实现分布式应用程序的创建。任何平台都有它的数据表示方法和类型系统。要实现互操作性,Web Service 平台必须提供一套标准的类型系统,用于沟通不同平台、编程语言和组件模型中的不同类型系统。目前这些协议主要有 3 种:①可扩展标记语言(XML);②概念 XML 结构定义(XSD);③简单对象访问协议(SOAP)。

Web Service 实现不同的系统之间能够用“软件-软件对话”的方式相互调用,打破了软件应用、网站和各种设备之间的格格不入的状态,完成了“基于 Web 无缝集成”的目标。

1.5.6 科学可视化技术

科学可视化^[22]是一个跨学科研究与应用领域,侧重于利用计算机图形学来

创建视觉图像,从而帮助人们理解那些采取错综复杂而又往往规模庞大的数字呈现形式的科学概念或结果,所以主要关注的是三维现象的可视化,如建筑学、气象学、医学或生物学方面的各种系统,重点在于对体、面以及光源等等的逼真渲染。它提供的是图形化表现形式的数值型数据,以便与科研工作者对这些数据进行定性和定量分析。

(1) OpenDX

OpenDX^[23]是一个 Linux 下的数据可视化工具,采用 C 语言开发。OpenDX 适用于科研、工程数据分析。OpenDX 原名为 IBM 的可视化分析软件——Data Explorer。在 1999 年 5 月 4 日,IBM 开放了该软件的源代码,并把它更名为 Open Visualization Data Explorer,简称 OpenDX。IBM Data Explorer 曾经是商业软件,但是由于这个软件在科研,教学中广泛运用后。IBM 改革了 Data Explorer 的版权方向于是出现了 IBM Visualization Data Explorer。由于使用版权的开放,学界将其称之为 OpenDX。

(2) 交互式数据语言(IDL)

IDL(Interactive Data Language)^[24]是一种数据分析和图像化应用程序及编程语言,先由美国 ITT 公司所有。

IDL 是完全面向矩阵的,因此它具有快速分析超大规模数据的能力。IDL 可以通过灵活方便的 I/O 分析任何数据。可以读取和输出任意有格式或者无格式的数据类型,支持通用文本及图形数据,并且支持在 NASA、TPT、NOAA 等机构中大量使用的 HDF、CDF 及 netCDF 等科学数据格式,以及医学扫描设备的 DICOM 标准格式。IDL 还支持字符、字节、16 位整型、长整型、浮点、双精度、复数等多种数据类型。从 IDL5.5 起还可以支持 MrSID 压缩数据格式。

IDL 支持 OpenGL 软件或硬件加速,可加速交互式的 2D 及 3D 数据分析、图像处理及可视化,除了保留传统的直接图形法外,IDL 还采用了先进的面向对象技术;可以实现曲面的旋转和飞行;用多光源进行阴影或照明处理;可观察实体(Volume)内部复杂的细节;一旦创建对象后,可从各个不同的视角对对象进行可视分析,而不用费时地反复重画。

IDL 具有强大的数据分析能力,从 IDL5.5 起支持多进程运算,IDL 带有完善的数学分析和统计软件包,提供强大的科学计算模型,支持 IMSL 函数库。它

的图像处理软件包提供了大量方便的分析工具、地图投影变换软件包,使您开发 GIS 易如反掌。IDL 提供了可缩放的 TrueType 字体,可以注记中文;能将结果存为标准图像格式或 PostScript 格式,并尽可能地使图像质量最优化。

IDL 支持 COM/ActiveX 组件,可将 IDL 应用开发集成到与 COM 兼容的环境中。从 Visual Basic、Visual C++等访问 IDL,还可以通过动态连接库和 COM 组件方式在 IDL 程序里调用 C、Fortran 等程序。用 IDL DataMiner 可快速访问、查询并管理与 ODBC 兼容的数据库,支持 Oracle, Informix, Sybase, MS SQL 等数据库。可以创建、删除、查询表格,执行任意的 SQL 命令,读取、设置、查询、增加、删除记录等操作。例如:图像和信号处理;3D、4D 曲线和曲面分析;三维体数据分析。

IDL 为用户提供了可视数据分析的解决方案,早在 1982 年 NASA 火星飞越航空器的开发就使用了 IDL 软件。IDL 使科学家无须写传统程序就可直接研究数据。IDL 还被广泛应用于地球科学、医学影像、图像处理、软件开发、大学教学、实验室研究、测试技术、天文、信号处理、防御工程、数学分析、统计等诸多领域。

(3) 矩阵实验室

MATLAB^[25]是由美国 mathworks 公司发布的主要面对科学计算、可视化以及交互式程序设计的高科技计算环境。它将数值分析、矩阵计算、科学数据可视化以及非线性动态系统的建模和仿真等诸多强大功能集成在一个易于使用的视窗环境中,为科学研究、工程设计以及必须进行有效数值计算的众多科学领域提供了一种全面的解决方案,并在很大程度上摆脱了传统非交互式程序设计语言(如 C、Fortran)的编辑模式,代表了当今国际科学计算软件的先进水平。

MATLAB 可以进行矩阵运算、绘制函数和数据、实现算法、创建用户界面、连接其他编程语言的程序等,主要应用于工程计算、控制设计、信号处理与通讯、图像处理、信号检测、金融建模设计与分析等领域。MATLAB 的基本数据单位是矩阵,它的指令表达式与数学、工程中常用的形式十分相似,故用 MATLAB 来解算问题要比用 C, FORTRAN 等语言完成相同的事情简捷得多,并且 MATLAB 也吸收了像 Maple 等软件的优点,使 MATLAB 成为一个强大的数学软件。

第二章 虚拟观测的研究进展

2.1 概述

卫星和望远镜观测精度的提高,使科学研究进入“数据雪崩”的时代。与此同时,数据网格、云计算等新技术的涌现,使得在协同环境下实现大规模分布式数据的共享成为可能。在这样的背景下,虚拟观测台应运而生,并迅速受到空间和天文学界的重视。

2002 年前后大批 VO 研究计划相继启动。这些 VO 项目按照学科可分为天文学和空间科学两类:在天文学领域,以美国国家虚拟天文台(NVO),欧洲虚拟天文台(EURO-VO),中国虚拟天文台(China-VO) 等为代表组成了国际虚拟天文台联盟^[26](IVOA);在空间科学领域,美国宇航局(NASA)的“Live With Sun”计划(VxO)按照“先学科后集成”的思路,提出了虚拟磁层观测台(VMO)、虚拟太阳观测台(VSO)、虚拟空间物理观测台(VSPO)、虚拟日地观测台(VHO)、虚拟宇宙线观测台(ViCRO)、虚拟电离层热层中间层观测台(VITMO)、虚拟辐射带观测台(ViBRO)、虚拟高能粒子观测台(VEPO)等八个子学科 VO,并通过相互间共享接口实现最终集成。

2.2 天文虚拟观测的研究现状

目前,国际虚拟天文台联盟成员已经发展为 19 个(见表 1),它们是美国虚拟天文台(NVO)、英国虚拟天文台(AstroGrid)、欧洲虚拟天文台(EURO—VO,其前身为目前已结题的天体物理虚拟天文台 AVO)、德国虚拟天文台(GAVO)、法国虚拟天文(OV—Prance)、中国虚拟天文台(China—VO)、欧空局虚拟天文台(ESA—VO)、日本虚拟天文台(JVO)、澳大利亚虚拟天文台(Aus—VO)、加拿大虚拟天文台(CVO)、意大利虚拟天文台(DRACO)、印度虚拟天文台(VO—India)、韩国虚拟天文台(KVO)、俄罗斯虚拟天文台(RVO)、匈牙利虚拟天文台(HVO)、亚美尼亚虚拟天文台(ArVO)、阿根廷虚拟天文台(AOVO)、巴西虚拟天文台(BRAVO)、欧盟南方虚拟天文台(ESOVO)等^[26]。

各成员联合成立了国际虚拟天文台联盟执行委员会,下设秘书和技术协调组,

按照侧重点的不同又细分为工作组和兴趣组。兴趣组主要是提出理论、应用和需求；工作组通过一系列的标准和原则来加以实现；技术协调组协调这两组的运行、意见和分歧。各组之间的协调配合使虚拟天文台项目更有效和健康有序地发展。

其中工作组又分为资源注册(Resource Registry)、数据模型(Data Modeling)、统一的内容描述(Unified Content Description)、数据获得层(Data Access Layer)、VO 事件(VO Event)、VO 表(VOTable)、VO 查询语言(VO Query Language)、标准和文档接受步骤(Standards Documents Acceptance Process)、网格和网页服务(Grid Web Services)；兴趣组分为 VO 的应用(VOApplications)、数据的保存和管理(Data Preservation and Curation)、VO 的理论(VO Theory)和热衷于全球网格论坛的天文研究组(GGF Astro RG)^[5]。

立项国家	项目名称
Australia	Australian Virtual Observatory (Aus-VO)
Canada	Canadian Virtual Observatory (CVO)
China	Chinese Virtual Observatory (China-VO)
Europe	Astrophysical Virtual Observatory (AVO)
German	German Astrophysical Virtual Observatory (GAVO)
India	Virtual Observatory India (VO-India)
Italy	Italian Data Grid for Astrophysical Research (IDGAR)
Japan	Japanese Virtual Observatory (JVO)
Korea	Korean Virtual Observatory (KVO)
Russia	Russian Virtual Observatory (RVO)
UK	Astrogrid
US	National Virtual Observatory (NVO)

表 1 虚拟观测项目一览表

2.2.1 英国天文网格

英国天文网格（AstroGrid）由英国粒子物理与天文研究委员会（PPARC）资助，是英国 e-Science 计划中网格计划之一。

英国天文网格的初期目标是一个具体的技术型目标，分为 7 点：①构建处理天文数据的数据网格；②开发高效的数据库查询和数据挖掘工具；③制定标准的数据库查询和数据挖掘的软件接口；④实现同时检索多个数据库的能力；⑤完成一套集成化的天文数据在线分析工具；⑥完成一套在线数据库分析和挖掘工具；⑦实现由用户上传算法并自动进行数据挖掘的机制；⑧开放式资源发现技术的研究^[5]。

随着项目的推进和与 IVOA 的融合，英国天文网格的科学家从科学目标和技术目标两个方面，修正了英国天文学界对 VO 的科学定位。其科学目标是：①改进在线天文研究的质量、效率、易用性、速度、投资回报率；②实现不同来源天文数据无缝透明的比较与融合；③消除跨学科研究工作中的数据分析障碍；④使数据密集型科研工作尽可能的容易开展并实现尽可能强大的功能。其技术目标是：①与 IVOA 伙伴一同开发被国际认可的数据、元数据、数据交换和数据履历（provenance）标准；为数据服务开发软件基础设施；②建立能为 AstroGrid 和主要数据中心共享的物理网格资源；建立和维护 AstroGrid 服务与资源注册；③基于英国主要数据库系统和天文学家的使用要求建立真正可用的虚拟天文台；④为 VO 系统提供用户界面；通过自主开发或者引进等方式为用户提供一套在 VO 环境下使用的科学工具；⑤为英国在 VO 领域树立领导地位。它的最终目标是基于英国国内的天文资源开发出真实可用并与网格体系兼容的 VO。

英国天文网格项目取得了巨大的成果。AstroGrid 通过多年的推进和执行，完成了全面的需求列表和科学问题文档，完成了科学需求制定。在体系结构方面，AstroGrid 从科学需求得到使用示例并由此确定 AstroGrid 的体系结构框架。在国际合作方面，AstroGrid 在世界范围内与 VO，网格，e-Science 领域的组织开展全面的合作，与 NVO，AVO 共同创立了 IVOA，并向 Globus 项目提交了 Community Authorization Service（CAS）的相关文档，还与 e-Science 其他计划合

作开展数据库领域的研究工作。在学科应用研究方面, AstroGrid 开展对 VO 相关技术的评测和调研, 完成了数个 AstroGrid 技术演示(如太阳、射电研究, 网格平台的建立等)与在线协同作业工具的开发, 还开发了一套功能全面的协作工具。

2.2.2 澳大利亚虚拟天文台

澳大利亚虚拟观测台项目^[5](Aus-VO)由澳大利亚望远镜国家设施(ATNF)组织发起。ATNF 隶属于联邦科学与工业研究组织(CSIRO), 负责澳大利亚射电天文学研究。CSIRO 是澳大利亚的国家科研组织。ATNF 管辖的望远镜有位于 Narrabri 的射电阵和 Parkes 以及 Mopra 的射电望远镜。这些望远镜都是 VLBI 的组成成员。澳大利亚虚拟观测台的数据资源主要来自澳大利亚射电阵公开的观测数据(ATCA)、HI Parkes 全天巡天数据(HIPASS)、悉尼大学 Molonglo 巡天数据(SUMSS)和英澳天文台数据(AAT)。

澳大利亚虚拟天文台的目标^[5]是, 作为国际虚拟天文台(IVO)的一部分, 实现澳大利亚天文数据和世界上主要天文台的数据的共享, 同时提供先进的科学工具, 使得澳大利亚的天文学家能作为 VO 前沿用户进行基础科学研究发展, 并在国际上继续巩固澳大利亚天文学的领先地位。

澳大利亚虚拟天文台项目执行至今, 主要取得了 6 个方面的成果: ①建立相关的数据标准、格式、压缩技术、协议, 实现 IVO 系统内的透明访问; ②为望远镜的观测数据开发了自动处理软件; ③建立全国性的分布式天文数据库, 并匹配计算能力; ④建立了连接国内外数据提供者与科研用户的高速网络; ⑤扩展交叉认证和可视化的能力; ⑥升级数据提供设施以保证提供高质量的数据。

2.2.3 欧洲天体物理虚拟天文台

欧洲天体物理虚拟天文台^[5](Astrophysical Virtual Observatory, AVO)由欧盟和六个欧洲组织联合资助。这六个合作组织分别是欧洲南方天文台(ESO)、欧洲太空局(ESA)、AstroGrid、CDS、法国 Louis Pasteur 大学、法国巴黎天体物理研究所 TERAPIX 天文数据中心、英国曼彻斯特大学 Jodrell Bank 天文台。

EURO-VO 提出了 3 个技术目标: ①在数据模型设计上, 侧重考虑 VO 的数

据兼容性；②在基础设施设计上，强调支持科学家在项目、讨论组等具体工作中对 VO 设施范围的资源和服务的高效利用；③建设一个能够支持更新优化 VO 组件、评鉴 VP 技术和设计实现新 VO 组件的新 VO。

为实现这一目标，项目首席科学家把 EURO-VO 的体系结构划分为 3 个模块，平行推进建设：

- (1) EURO-VO 数据中心：数据中心是一个欧盟数据中心联盟，数据中心的功能是，发布数据资源，提供数据资源的物理存储；
- (2) EURO-VO 设施中心：设施中心是一个为 VO 提供集中式的资源注册、数据模型和标准、认证机制和 VO 技术的社区支持；
- (3) EURO-VO 技术中心：技术中心是一个分布式的组织，它可以协调学科工具、VO 系统的工作流，使之服务于研究。

2.2.4 日本虚拟天文台

日本虚拟天文台^[5]（JVO）由日本国立天文台（NAOJ）的天文数据分析中心牵头，于 2002 年 4 月被确立为日本国立天文台的正式研究项目。目前，JVO 的主要参建单位为日本国立天文台、Ochanomizu 大学和富士通公司。

经过一年多的努力，JVO 已经取得了让人瞩目的成果。其中最重要的是开发了日本虚拟天文台查询语言（JVO QL），取得了虚拟天文台查询语言（VOQL）研发的领导权。JVOQL 是 JVO 在 2002 年定义的一个统一的查询语言，用以实现对天文数据库的访问。作为 SQL 的一种扩展，JVOQL 可以实现对存储于数据库中的星表和图像数据的查询检索。JVO 努力通过单一的用户接口实现对多个数据服务器上星表数据和图像数据的检索提取功能。

JVO 利用网格技术建立了一个原型平台来验证 JVOQL 对多数据库查询的支持。这个平台利用 Java 实现用户接口，利用 GT2 进行相关服务开发，以便可以与 OGSA 兼容。JVO 的大多数功能都是利用自由软件实现的。JVO 计划建立自己的数据库系统，把从远程服务器上提取的数据存储到本地数据库中。

JVO 目前的工作主要集中在三方面：①OGSA 的广泛应用；②数据分析工具的开发；③用户访问控制。

2.2.5 美国国家虚拟天文台

美国国家虚拟天文台^[5] (NVO) 是基于天文数据库和网络信息技术, 利用强大的计算机与网络的软硬件支持构件的虚拟观测系统。NVO 将来自地面和空间观测站的天文观测数据有机地统一在一起, 目标是最大化来自这些观测数据的科学潜力, 将数据以统一标准的格式提供给专业研究者、天文爱好者和学生。

NVO 项目成员来自著名的天文台、大学、国家实验室和公司, 包括: 约翰霍布金斯大学, 加州理工学院, 美国国家光学天文台, 太空望远镜科学研究所, 加利福尼亚大学, 宾夕法尼亚大学, 费米国家实验室, NASA, 施密松天体物理天文台, 卡耐基梅隆大学, 美国海军天文台, 伊利诺斯大学, 匹兹堡大学, Lawrence Livermore 国家实验室, 南加州大学, 威斯康辛大学, 芝加哥大学, 微软研究院, 圣地亚哥超级计算中心等单位。此外还包括欧洲空间局, 加拿大天文数据中心, CDS, 爱丁堡皇家天文台, 贝尔法斯特大学, 澳大利亚望远镜国家设施 (ATNF), 东京大学, 欧洲南方天文台等国际伙伴。

目前, NVO 的主要科学和技术成就反映在标准协议制定和软件应用开发两方面。在标准协议制定方面, NVO 提出了锥形检索 (ConeSearch); 简单图像访问协议 (SIAP); VOTable; 统一内容描述标准 (UCD) 等核心标准。在软件应用开发方面, NVO 硕果累累, 相继研发出如下学科应用: ①VOTable 解析; ②JAVOT: VOTable 的 Java 语言解析器; ③SAVOT: VOTable 简单访问程序; ④VOTable 浏览器; ⑤VOPlot: VOTable 数据画图程序; ⑥VOTool: VOTable 数据可视化与编辑程序; ⑥Treeview: 天文资源浏览程序; ⑦ ConeSearch 注册等。此外, NVO 对 VO 概念的宣传、推广做了大量的工作, 作为创始成员与 AVO、AstroGrid 一同创立了 IVOA, 并承担了 IVOA 社区内部许多协调管理工作。

在 2003 年的美国天文学会 (AAS) 201 次会议上, NVO 成功的展示了自己的三个应用演示。

(1) γ 射线暴事件追踪

利用远程服务提供的标准协议, 比如 ConeSearch、SIAP、VOTable, 和标准的语义编码 (比如 UCD) 来进行数据提取和转换。在多个波段上收集关于某个天区或者有趣的瞬间事件, 比如 γ 射线暴, 的图像、星表等各种相关数据。这个 demo 重点测试了 VOTable 的使用。

(2) 褐矮星候选体搜寻

对 SDSS EDR 光学巡天星表和 2MASS 红外巡天点源星表进行交叉证认来搜寻褐矮星的候选体。这个 demo 重点测试了 NVO 对 ConeSearch 以及交叉证认功能的支持。

(3) 星系形态学分析

利用 VO 提供的通过标准接口对分布式异构星表和图像数据的访问能力、建立在标准数据交换格式 (FITS、VOTable) 上的数据整合能力和网格环境下的高性能计算能力, 对星系的表面亮度、集中指数 (Concentration Index) 和不对称指数 (Asymmetry Index) 进行了计算, 并利用这些计算得到的参数与其他数据, 比如测光数据、本动速度、团中位置、星系团大尺度结构, 进行联合分析, 来进行星系团动力学研究和大尺度结构下的星系演化研究。

2.2.6 中国虚拟天文台

中科院国家天文台的虚拟天文台 (China-VO) 建设思路是: 把国内天基或者地基的望远镜拍摄数据组织归档, 形成虚拟星空, 并利用数据库检索技术把用户需求转化为特定观测区域的方式建设虚拟望远镜, 虚拟星空和虚拟望远镜基础上, 提供图像处理、数据分析、文献检索服务, 构造虚拟天文学社区。

中国虚拟天文台项目主管崔振洲博士认为, 虚拟天文台是通过先进的信息技术将全球范围内的研究资源无缝透明连接在一起形成的数据密集型网络化天文研究与科普教育平台。虚拟天文台是技术使能的, 是基于先进的 IT 技术来实现。虚拟天文台利用的是全球范围内的资源, 不仅局限于一个大学、一个天文台或者一个局域网。虚拟天文台中的资源是无缝透明地融合在一起的, 用户不会感觉到它们在物理位置上的差异。还有, 虚拟天文台是数据密集型的, 它的目的是要让所有能够接触到互联网的人都可以得到真实而丰富的天文数据, 进而充分发挥这些数据的科学价值。

当前国家天文台的虚拟天文台建设主要集中于 5 个部分的工作: ①中国虚拟天文台系统平台的开发; ②国内外天文研究资源的统一访问; ③支持 VO 的项目与观测设施; ④基于 VO 的天文研究示范; ⑤基于 VO 的天文科普教育。

2.2.7 加拿大虚拟天文台

加拿大虚拟天文台^[5] (CVO) 的基本设计目标是为科学家开发一套系统。这个系统将不仅仅能让天文学家找到数据,还必须能产出科学的结果。这些结果是通过其他途径无法或很难完成的。从技术的角度看, CVO 将利用最好的技术和实践经验来实现自己的目标。CVO 将不被“向前兼容”问题所困扰,不会将以前加拿大天文数据中心 (CADC) 的服务全部囊括到新的系统中。

CVO 的基本设计目标是实现全波段的科学研究。他们的设计思路是:通过抽象,将技术细节和科学分离,将技术隐藏起来从而实现对数据的统一访问;通过定义一个足够通用的数据模型来满足所有天文学家的需要。他们认识到不同波段研究工作的技术细节很不相同,可以开发一个系统对 VO 中不同类型的内容进行抽象从而去掉这些技术细节和复杂性。

目前, CADC 设计了一个通用的科学数据模型和一个访问科学数据仓库的 API。他们正利用 Java 程序设计语言开发一系列的 Jini 服务,搭建 CVO 的原型。

此外,加拿大天文数据中心决定发展从大的科学数据库中进行数据挖掘的工具,提出发展数据挖掘技术的提案。CADC 在过去十年在数据存储、加工、存档、分布等方面处于领先地位,是首家使用 WWW 界面服务器,同时又是第一个使用 CD-ROM 技术的场所,而且与国际上许多单位都有过成功的合作关系,在计算机软件方面成果斐然。

在各国争相出笼虚拟天文台的形势下,加拿大天文学界为保持其在国际方面的领先优势,增加该国天文学家的工作效率,提出了在加拿大天文数据中心发展数据挖掘中心的提案,为逐步过渡到全球性虚拟天文台做准备。

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

2.3 空间科学虚拟观测的研究现状

2001 年, NASA 在国际大地测量学与地球物理学联合会中, 提出把发展地学空间科学的虚拟观测台作为核心任务之一。2004 年 10 月, NASA 以白皮书方式提出“空间与太阳物理虚拟观测台框架”^[27], 在此框架下按照“先科学后集成”的机制推动了多个虚拟观测台发展计划。2007 年, NASA 将发展基于虚拟观测的数据应用环境列为其日地空间物理学科长远发展纲要的内容之一。同时, NASA 在“与日同在计划”(Live With Sun)首先提出结合该项目优先建设“一体化的日地空间数据应用环境(HPDE)”, 即面向子学科的日地空间虚拟观测台(Sun-Earth Connections VxOs), 这些 VxO 分别是:

- 虚拟太阳观测台(VSO);
- 虚拟空间物理观测台(VSPO);
- 虚拟太阳物理观测台(VHO);
- 虚拟宇宙线观测站(ViCRO);
- 虚拟磁层观测台(VMO);
- 虚拟电离层、热层、中间层观测台(VITMO);
- 虚拟辐射带观测台(ViRBO);
- 虚拟高能粒子观测台(VEPO)。

2.3.1 虚拟太阳观测台(VSO)

一直以来, 物理学家就太阳物理数据集研究提出应该发展一个系统, 以使得数据分析和研究更高效。虚拟空间物理观测台是已有工作与该观点的继承和发展而成。因而, 虚拟太阳观测台^[28]开发的目的在于, 让太阳物理学家在使用多个独立数据集(3-10 个以上)进行物理研究时更高效快速。通常地, 针对每一个物理数据集, 研究人员必须能够:

- (1) 定位数据在虚拟观测台的位置;
- (2) 通过数据资源提供者的访问接口检索数据;
- (3) 查找并选择相关数据;
- (4) 快速高效的传输物理数据。

然而, 以上每一步都需要大量的耗时, 而且每一步都没有形成一个顺畅的工

作流,这已经成为当今科学研究的限制瓶颈。虚拟太阳观测台针对该瓶颈,结合 IT 技术把 4 大步骤进行整合,形成科学工作流,并集成到统一的系统中,使得研究人员能够高效的查找到他们所需要的数据资源。

因而,虚拟太阳观测台开启了一个新的大规模数据资源协同工作的太阳研究方式。而这种研究方式在过去是从来没有过的。目前,虚拟太阳观测台的建设主要为实现 5 个科学目的^[29]:①提高在太阳物理领域数据归档的质量;②提出并发展一套囊括数据、技术和管理的系统,向数据资源提供者提供简单便捷的共享数据的方法;③提供分析工具以提高数据挖掘和基于内容检索的效率;④虚拟太阳观测台是一个允许用户快速高效地检索和获得符合学科需求的数据产品的中间件集合。

依据 NASA 对虚拟太阳观测台的定位^[30],VSO 必须包含两个基本功能:检索数据产品和传输数据产品。与此同时,虚拟太阳观测台还要向太阳物理研究社区提供更多的满足学科需求的用户接口。这不但包括系统内的接口,还会考虑 VSO 外的用户的使用接口。因而,VSO 必须具有较好的开放性。因为用户接口可能是基于 web,或是基于命令行或是基于桌面应用,或者是基于其他嵌入式的系统平台。无论是基于哪个异构的平台,在设计上 VSO 将会遵循如下 workflow 模式:①用户接口向用户展现检索选项;②用户发送检索条件给用户接口;③用户接口向 VSO 请求符合检索条件的数据资源;④VSO 向用户接口返回合适的数据集;⑤用户接口向用户展现数据检索结果;⑥用户通知用户接口感兴趣的数据;⑦用户接口向 VSO 通过特定的数据传输模式请求获取数据产品;⑧VSO 按照指定数据传输模式向用户接口返回数据产品反馈结果;⑨用户接口依据反馈结果联系数据资源提供者并获得数据资源。

2.3.2 虚拟空间物理观测台(VSPO)

虚拟空间物理观测台^[31]提供一个访问数据产品、模型产品以及其他类型的空间物理数据产品和对应元数据的途径,同时它拥有传输与共享数据产品的能力。VSPO 具备 3 大亮点:①多维度的检索机制;②丰富的模式计算;③完备的数据分类体系。

目前,虚拟空间物理观测台具有多类型多维度的检索机制,兼容模糊检索、

时间检索和元素检索，同时也融合模型了大量的计算模型，主要包括：离子体俘获粒子模型（AE-8、AP-8）、CCMC 磁场模型、电离层/大气模型、地球模型（heliospheric models）、太阳模型（solar models）、CIRA-86 大气模型、国际地磁场参考模型、MSIS 大气模型等，部分模型更能提供在线的模型可视化。VSPO 整合 ACE、WIND 等多类卫星的轨道数据，部分轨道模型可以直接在线进行计算演示。

VSPO 具备完整的数据分类体系。从数据类型上看，数据检索可以按 catalog、display (image/movie)、display (plot)、model、numerical data 进行粗粒度分类后，再依据 binary、cdf、fits、ctif、hdf5、html、jpeg、mpeg、pdf、png、tiff、text 等进行细粒度分类。从空间事件上看，VSPO 的编目数据提供了基于 STEREO、IMAGE、SOHO 等卫星监测到的 FUV 亚暴、弓形基波穿越、行星际扰动、CME 等事件的划分。从观测区域上看，VSPO 还按照空间分区进行数据整理，它能够按照地球磁鞘、地球磁气圈、地球磁气圈磁尾、地球磁气圈干线（main）、地球磁气圈极面、地球磁气圈辐射带、地球近表面、地球近表面大旗、地球近表面电离层、地球表面、日光层、太阳等进行划分。从光谱范围角度看，VSPO 的数据又可以划分为：远紫外区、红外线、光学、无线电频率、紫外线、x 射线。从测量类型角度看，VSPO 的数据还可以划分为：activity index 指数，电场、高能粒子、图像强度、粒子成分、光辉、磁场、剖面、频谱、辐射率、热等离子体、粕厂、多普勒、星历表、磁力图。

2.3.3 虚拟日地观测台(VHO)

虚拟日地观测台^[32]的设计思路体现在着重强调数据检索和发现、学科应用服务和数据产品处理工具。为提高虚拟日地观测台的数据兼容性和融合性，VHO 提供从 VxO 中导入数据的方式。

2.3.4 虚拟宇宙线观测站(ViCRO)

虚拟宇宙射线观测台^[33]的提出是针对已存在或者正在发展的日地虚拟观测台，面向特别重要的宇宙射线数据集而提出的一个扩大处理能力的解决方案。虚拟宇宙射线观测台，主要把数据处理和共享的关注点放在太阳和日地空间的学

科应用以及数据之间的互操作。虚拟宇宙射线观测台的核心操作数据源主要包括：Heliospheric Network (HN)飞船的宇宙射线科学仪器，HN 飞船的星历数据服务，太阳系物质，地基的中子监视器和高纬度热气球设备的能宇宙射线探测，虚拟宇宙射线观测台通过 VHO 项目开发的通用中间件接口，支持数据查找，访问和其他 VxO 的接口。

2.3.5 虚拟磁层观测台(VMO)

虚拟磁层观测台^[34]是由马里兰大学的 Jan Merka 教授为 PI 的一个项目组所进行的项目，该项目的承建单位主要包括：NASA，University of Maryland，University of Boston，University of Colorado，Hampton University，University of California Berkeley，NOAA，Johns Hopkins University，外加两个专业公司：ScriberQuest Inc. 和 L-3 Com. GSI。

虚拟磁层观测台设立的主要目的在于提高磁层数据集的数据类型发现，比如哪一类数据可用，宇宙飞船或者观测设备；智能化的内容检索，能够组合成复杂的查询条件以快速检索所需数据；多飞船探测数据研究，比如 dayside 磁层的相互作用，亚暴的相互作用和关联分析，磁场重联分析。

虚拟磁层观测台的主要任务是帮助磁层研究的科学家高效的发现和使用磁层数据。为了保证 VMO 项目的建设更贴近科学家的需求，VMO 项目组开展了一系列的论证和评论会议，论证结果如下：在 VMO 的用户接口和 API 的设计方面，必须考虑系统的特色应用和可用性，并收集用户的访问和使用情况统计数据；在 VMO 的设计方面，重用当前已经部署的 VxO 中成功的组件和特色应用，并且能够和 VxO 进行互操作。

VMO 认为所有使用 VMO 的 API 开发的 web 或者本地的应用都称为一个网格服务。VMO 的核心服务是基于用户提供的查询，在磁层数据网格中定位和传输符合用户需求的数据指针。增值服务是，提供模块化的社区需求驱动的支持互操作的 VMO API，比如：①通过数据挖掘技术提供更智能化的数据检索能力；②通过 PAPCO 提供强大的数据虚拟化和分析服务；③通过 ViSBARD 对飞船轨道进行 3D 展示；④通过提供 VxO 的接口库。

2.3.6 虚拟电离层、热层、中间层观测台(VITMO)

虚拟电离层热层中间层观测台^[35]是一个着重于数据发现的网格系统。它允许用户通过统一简易的接口发现网格中所有的数据,包括:新加入和已存在的电离层热层中间层数据。

VITMO 主要面向研究电离层、热层和中间层的科学家,VITMO 项目由 Johns Hopkins University Applied Physics Laboratory 牵头实施,联合 NASA Goddard Space Flight Center, Hampton University, University of Michigan, University of Texas at Dallas, and High Altitude Observatory/National Center for Atmospheric Research 等 5 家单位承建,主要提供天基和地基的电离层、热层、中间层的学科服务如:数据、工具、模型的集成接口。

VITMO 的数据主要来自地基的遥感设备和卫星遥感设备。另外,从数据物理内在驱动关系审视,太阳辐射、太阳风和磁场粒子之间的相互关系,是有内在的物理驱动因素,因而 VITMO 项目设立的目的是处理涵盖电离层、热层和中间层的大规模、大范围的观测数据,更重要的是如何在海量的观测数据里发掘物理驱动因素,显得非常重要和迫切。

VITMO 的特色主要包含如下 5 点:①是一个基于关键词的检索接口,检索接口重点涵盖电离层、热层和中间层的物理现象、基础物理性质和通用数据测量标称的专业术语;②提高检索服务的质量,提供卫星和地面位置的联合检索,并可以通过地球物理条件限制搜索条件;③提供单一数据传输机制,该传输机制支持各类观测数据,包括:数据和服务的访问指针;④提供基于图片型数据概要描述的检索;⑤数据提供者的网站与数据浏览过程相集成。

2.3.7 虚拟辐射带天文台(ViRBO)

虚拟辐射带观测台^[36]ViRBO (Virtual Radiation Belt Observatory),是一个面向专属研究领域的虚拟观测台。ViRBO 项目开始于 2006 年秋,由 NASA Heliophysics Data Environment Program 资助。ViRBO 提供地球辐射带的相关数据和学科服务的访问接口,作为 ViRBO 的项目开发的一部分,承建单位开发或者扩展了一系列的现存学科软件库(Autoplot, VxOware)。这些学科软件库,不仅可以适用于地球辐射带学科领域,同时也适用于 VxO 中的其他学科。

ViRBO 建立的目的是为研究员，实验师和工程师提供一个统一的平台。在该平台上不仅能够获得高质量的校正的辐射带数据和模型输出，还集成必要的分析工具；发展一个通用的可扩展的开源的虚拟观测台系统，能够允许新的 VxO 项目基于该系统进行开发，从而加快 VxO 框架的建立。

项目参加人员	
PI	Bob Weigel (George Mason University)
Software	Eric Kihn (NOAA/NGDC, ViRBO Web and API)
Software	Mikhail Zhizhin (RFO, ViRBO Web and API)
Software	Jeremy Fadden (Cottage Systems, ViRBO Software)
Data	Dan Baker (CU/LASP, Data and metadata)
Data	Sebastian Bourdarie (Onera, Data calibration)
Data	Janet Green (NOAA/SEC, Data and metadata)
Data	Shri Kanekal (CU/LASP, Data and metadata)
Data	Reiner Friedel (LANL, Data and calibration)
Data	Paul O' Brien (Aerospace, Data and metadata)

表 2 ViRBO 参建单位

2.3.8 虚拟高能粒子观测台(VEPO)

VEPO^[37]项目的领导者是美国宇航局(NASA)戈达德太空飞行中心(Goddard Space Flight Center)的太阳物理学科学部下辖的日光层物理实验室。虚拟高能粒子观测台(VEPO)服务于太阳物理学数据用户社区。

VEPO 作为一个 VHO 核心组件，其关注范围在虚拟日地观测台之内。VEPO 的目的在于整合和提高来自航天器和亚轨道观测设备的高能粒子数据产品的发现、获取、分析和使用的效率。VEPO 主要关注的是来自日地空间物理研究领域的科学家，目前 VEPO 小组正致力于进行长远的规划，通过整理高能粒子的仪器字典，对关键数据产品进行描述和注册，并引入数据查询的 VHO 接口访问，从而提高 VEPO 和 VxO 兼容性。VEPO 将会引入空间物理档案检索与抽取语言，改进来自太阳风层探测器的高能粒子数据产品的适用性。

VEPO 整合戈达德太空飞行中心的 Goddard's Space Physics Data Facility (SPDF), 以及 ACE Science Center (ASC)、the Stereo Science Center (SSC)、the Ulysses Data System (UDS)、the multi-mission on-line archives of Fundamental Technologies、LLC 的数据源服务。

第三章 虚拟观测台体系结构的分类研究与典例分析

3.1 概述

要构建一个 VO，其核心是体系结构设计。随着 VO 体系结构研究的深入，研究人员发现，紧密结合学科数据应用特点，作针对性的体系结构设计才是 VO 发展的关键。本章将在国内外空间科学虚拟观测台体系结构研究成果的基础上，先提出空间科学虚拟观测台体系结构的一个分类方法，然后介绍典型空间科学虚拟观测台的体系结构的研究结论。

3.2 空间科学虚拟观测台的体系结构分类

目前，按照体系结构，VO 可分为 3 类：①Query-Triggered 型体系结构，以查询作为科研的触发点，重点关注资源的检索发现，以 VMO 为典型。②Application-Triggered 型体系结构则认为学科应用才是科研的起点，数据发现、访问、传输等功能被封装在学科应用中，以印度虚拟天文台(India VO)为典型。③Hybrid 型体系结构，是当前主流的体系结构，通常采用面向服务的思想，服务与服务之间通过数据管道进行 IO 操作，以 NVO 为典型。

3.3 VSO 体系结构分析

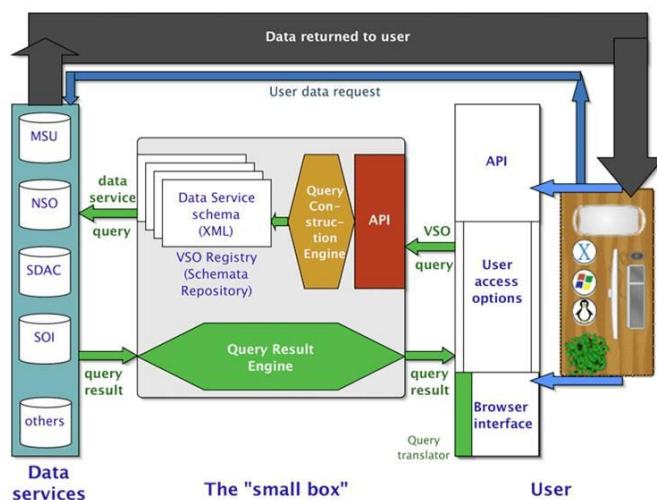


图 3-1 虚拟太阳观测台体系结构

虚拟太阳观测台是 Query-Triggered 型的体系结构（见图 3-1）。VSO 的体系结构自下而上可以分为三层：

(1) 数据层

数据层是一个多节点的分布式数据系统。数据层对外提供两类接口。第一类是分布式数据查询接口；第二类是数据结果返回接口。

(2) 数据发现服务层

数据发现服务层是一个集中式的数据检索系统，即系统只有一个资源注册器，以存储元数据。该资源注册器自动与分布在各地的节点进行元数据收割和同步，以保证资源的时效性和全局发现。用户的查询通过 API 提交到查询构造引擎后，被转化为底层查询原语提交到 VSO 的资源注册器。VSO 的资源注册器执行查询原语后，把查询结果交给查询结果引擎。查询结果引擎完成解释、排序和展现后返回给用户。

(3) 用户接口层

该层主要提供与用户进行交互的桥梁和门户，同时支持不同桌面应用的再开发。

在 VSO 体系结构进行原型设计过程中，设计人员提出 VSO 体系结构应具有面向如下几个特征去设计：①面向太阳数据的管理模型，实现分布式数据检索、定位、查询优化的虚拟化；②支持 BS 体系和 CS 体系的应用开发接口；③ 充分兼容和利用现有的数据服务，而不是重新设计元数据标准；④数据访问绕过 VSO 作为独立部件，科研工作者不需要直接以 VSO 作为桥梁来访问数据；④必须要易于在分布式系统中，动态扩展数据资源。

虚拟太阳观测台的体系结构设计致力于通过在已有的功能上增加抽象层，尽可能减少数据资源提供者的负担，即允许数据资源提供者继续使用其现存的传输机制和数据目录，却不需要依据新的数据模型对数据资源进行整理，也不需资源提供者必须支持一个或许并不能很好吻合其数据管理模式的数据转换机制。

为实现如上的目标，VSO 提出了体系结构设计的几个基本原则：①系统设计应能自动适应公共数据模型的调整；②系统设计应能保持最低的数据资源维护成本；③系统设计应尽可能减低请求加入数据网络的资源提供者的系统改变和维护负担；④系统设计应具有可扩展性，允许资源提供者在提供数据资源之余，也

可以向系统提供额外的服务；⑤系统设计应能使得太阳物理研究人员工作更有效率和效果。

3.4 VMO 体系结构分析

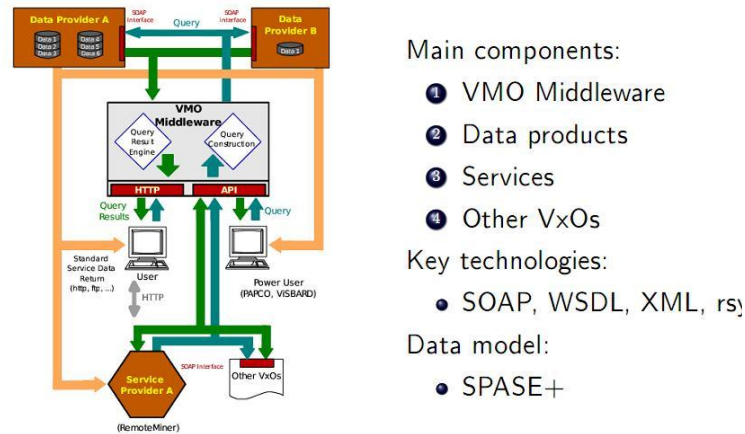


图 3-2 虚拟磁层观测台体系结构

VMO 属于 Query-Triggered 型体系结构（见图 3-2）。其体系结构可以划分为四层：数据层、中间件层、用户层、应用服务层。VMO 的数据层是一个分布式的数据系统。但是和 VSO 不同，VMO 的认为数据系统的基本单位是数据提供者，每一个数据提供者可以拥有一个或者多个数据节点。中间件层的核心是一个数据搜索引擎，主要分为两大部分：查询构建与解释引擎和查询结果引擎。与 VSO 不同的是，VMO 属于分布式查询，而不是集中式查询，即每一个数据提供者都有自己的 Database Management System。当中间件层接收到来自用户的数据查询请求，数据搜索引擎会基于元数据模型，构建符合语法与语义的查询原语，把原语分发给不同的数据提供者。数据提供者接收到来自中间件层的原语后，各自检索自己的资源注册器中的元数据，并把查询结果返回给中间件层。由中间件层进行结果合并，最终提交给用户。

在体系结构上，分布式查询与集中式查询的区别在于，分布式查询具有更低的耦合性。资源发现和资源存储、访问、加入、撤销等彻底分离。这种低耦合的体系结构设计，易于提高系统的透明性和健壮性，但是却难于实现系统的一致性和实时性。

VMO 体系结构一个重要理念是把学科应用服务和面向基础结构功能的服务

分离出来，作为一个独立的层次，进行重新规划和设计。VMO 认为，学科应用服务才是一个虚拟观测系统的价值所在，不仅如此，VMO 还非常重视学科应用于其他 VxO 数据访问接口的兼容性。

3.5 VITMO 体系结构分析

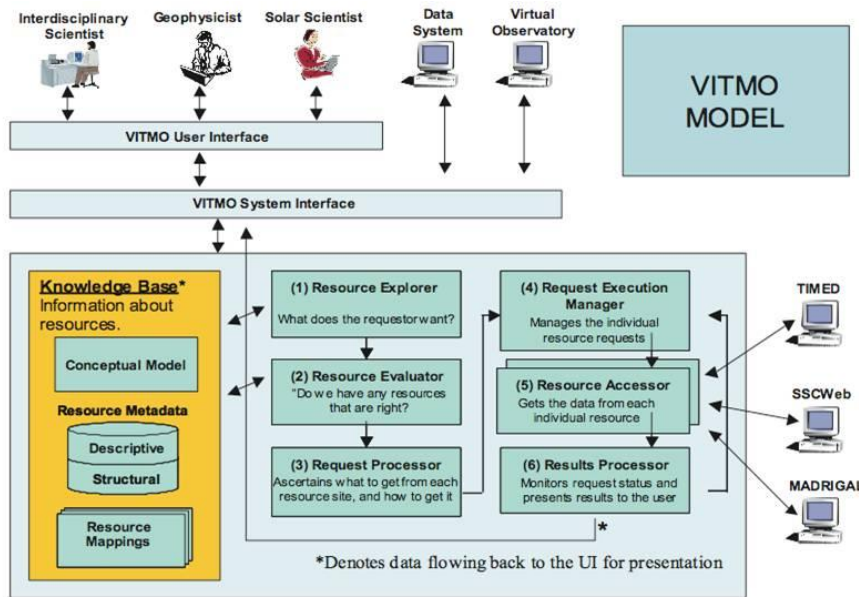


图 3-3 虚拟电离层、热层、中间层观测台体系结构

VITMO 的体系结构属于 Query-Triggered 型体系结构（见图 3-3）。其体系结构可以分为三个层次：资源层、资源发现层、接口层。

VITMO 资源层主要由两部分组成：元数据和物理观测数据。元数据是按照学科数据标准，对物理观测数据资源的描述，是数据的数据，其目的是实现资源快速高效的发现；物理观测数据，是卫星载荷与地面台站针对特定物理要素进行观测的数据记录，是科学价值的最终载体，是科学应用与产出的根本。

与 VMO 和 VSO 的资源发现层不同，VITMO 对资源发现层系进行了更详细的设计，按照科学工作流，分为 6 个部分：资源浏览，用以解释科研用户的学科数据需求；资源评估，用以分析资源层中是否具有科研用户请求的资源；查询处理，把高级查询转化为低级原语查询；查询执行，由系统执行查询原语，返回查询结果；资源访问，系统基于查询结果，通过特定访问协议访问并获取数据资源；结果处理，系统把查询结果以及相关数据资源传输给科研用户。

VITMO 接口层主要包括系统开发接口和用户访问接口。其中，系统开发接口的目的是，为日后的学科应用设计、桌面软件设计以及与其他 VxO 进行数据交互提供可能性；而用户访问接口则是一个基于 web2.0 技术的门户。

3.6 VSPO 体系结构分析

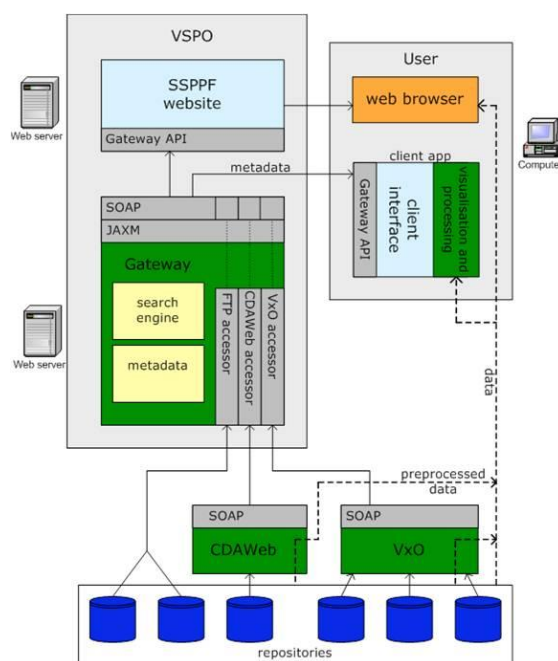


图 3-4 虚拟空间物理观测台体系结构

VSPO 的体系结构是一个典型的 Hybrid 型体系结构（见图 3-4），由三个核心要素组成，分别是：资源、VSPO 和用户。每一个要素，独立构成一个独特的软件体系结构。其中，资源的体系结构可以分为三层：物理数据层、元数据层、访问接口层。值得一提的是，资源要素的接口层拥有 CDAWeb、VxO 接口层和 SOAP 接口层等三类接口，能够支持科研用户通过 CDAWeb 接口直接访问物理数据、其他虚拟观测系统通过 VxO 系统访问物理数据、VSPO 搜索引擎通过 SOAP 接口访问元数据。

VSPO 由接口、搜索引擎和数据访问器三部分组成。接口模块既可以通过 SPPF Web API 直接向用户提供服务，也可以允许 VxO 或者学科应用通过 Gateway API 和 SOAP 接口，以 VSPO 为中介，查询和访问数据。与传统数据应用系统一样，VSPO 也是采用在元数据的基础上，建立搜索引擎的方式实现数据

的检索发现。但是，VSPO 本身并不负责数据资源层的建设，只是通过设计多协议的数据访问器（FTP 协议、CDAWeb 访问协议、VxO API 等）访问科学数据，从而实现了功能和数据的分离。用户层则包括基于浏览器的 web 访问接口、基于桌面应用的程序访问接口以及基于 VxO 的用户访问接口。

3.7 VHO 体系结构分析

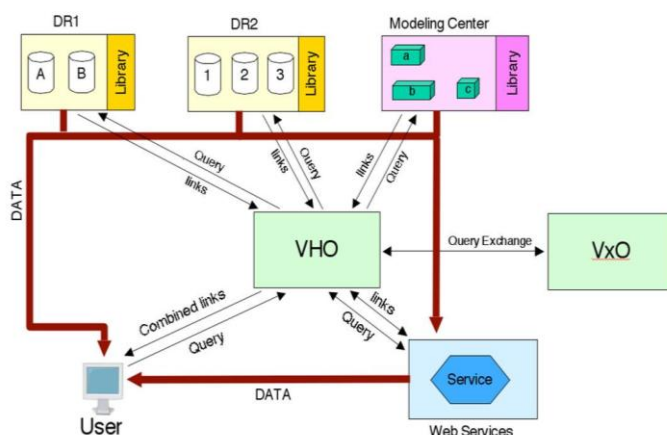


图 3-5 虚拟日地观测台体系结构

VHO 属于 Hybrid 型体系结构中的一种。VHO 的体系结构设计（见图 3-5）的耦合度非常低，以至于 VHO 对于科研用户而言，仅仅是访问接口和功能汇聚网关。VHO 认为一个完整的虚拟观测系统，应该包含数据资源、用户、模式计算中心、支撑数据应用环境的基础服务以及规范开放的接口，而 VHO 则仅仅面向用户请求，充当“路由”角色，实现访问请求和请求结果的转换、传递和展现。相比起 Query-Triggered 型的虚拟观测系统，VHO 更重视虚拟观测系统的计算和应用功能，并把模式计算和应用作为系统的一个独立的核心组成要素。

3.8 IVOA 体系结构分析

IVOA 的最新报告分成三个级别（0 级、1 级和 2 级）^[38]，从粗粒度到细粒度渐进地讨论体系结构的设计。

天文学专家对 LEVEL-0 级体系结构（图 3-6）作如下定义，系统结构的描述粒度仅局限于，划分层次和定义每一层次的基本功能。Level 0 级别的天文 VO

的体系结构被分为三层：用户层、基础设施层和资源层。其中，用户层主要包含基于浏览器、桌面和脚本的应用；基础设施层主要指资源注册、数据访问和 VO 核心；资源层是一个数据和元数据的集合。

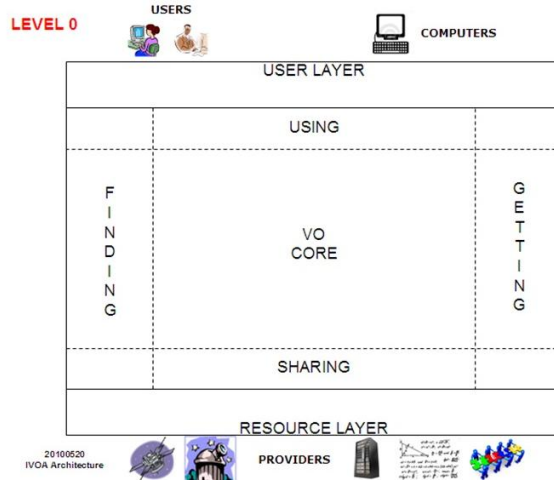


图 3-6 IVOA 虚拟观测台 0 级体系结构

图 3-7 展示的 LEVEL-1 级体系结构，是基于 LEVEL-0 级体系结构，进一步指出每一层的服务或数据模型，并分析层次之间通信交互的方式。

LEVEL-2 级体系结构（见图 3-7），则在 LEVEL-1 级体系结构的基础上，为每一个服务的功能点进行具体划分，并明确实现该功能点的技术路线、实现思路和应用方法。

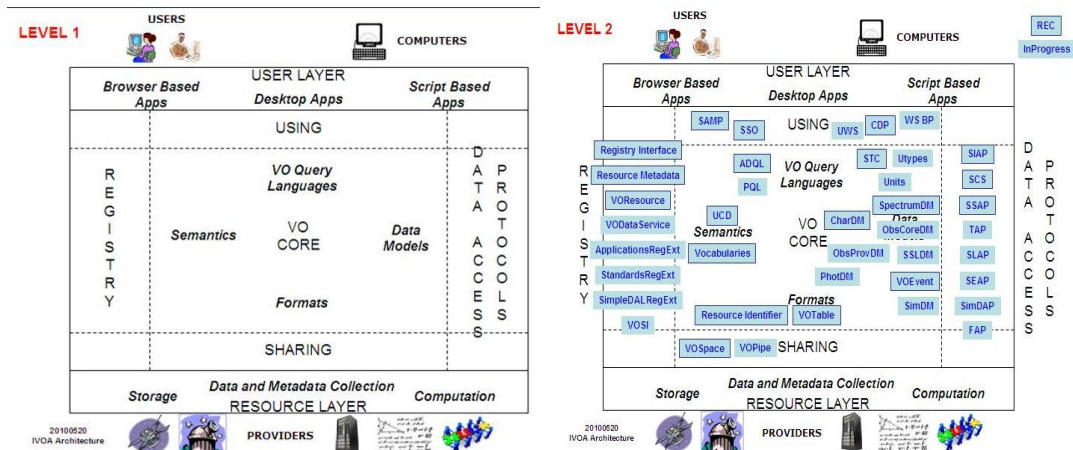


图 3-7 IVOA 虚拟观测台 1 级和 2 级体系结构

第四章 空间科学虚拟观测台体系结构的设计理论研究

4.1 概述

为此,在天文领域,IVOA 专门针对天文主流数据形式“星表”,制定基于 XML 的表列数据集合规范(VOTable),以满足数据互操作的需求。VOTable 包含星表的表数据和元数据,不仅支持与应用软件进行互操作,还可以基于 VO Query Language(VOQL)进行查询。因而,利用 VOTable,天文 VO 更容易实现数据的一致访问以及学科应用的集成。

与天文数据不同,空间科学数据具有以下特点:①只有行星科学的 PDS 标准和日地空间物理的 SPASE 标准,缺少统一的数据管理模型。②以文件卷宗为数据归档的主要组织方式,难以构建类 VOQL 的学科查询语言。③数据格式复杂多样(CDF、netCDF、DAT、HDF、CDFML 等),且无主流格式。因而,在数据管道的过程中,往往需要引入阅读和格式转换工具。④不同观测对象的数据之间存在相关性,如日冕物质抛射与磁场扰动的时序因果关系。

为适应空间科学数据的特点,我们有必要面向行星科学和日地空间物理两大学科,研究如何设计空间科学 VO 的体系结构。

4.2 核心问题分析

空间科学虚拟观测台的目标是实现数据资源、计算资源、存储资源和知识资源的无缝链接与全面共享,为空间科学研究提供虚拟化的应用集成环境。为此,我们在体系结构设计中着力解决如下七个问题:

(1) 行星科学和日地空间物理的数据管理模型异构

PDS 和 SPASE 分属行星科学和日地空间物理学。研究对象的差异带来两者在属性选取上的区别。因而,要融合两个学科数据,就必须实现两个学科数据标准的统一。为此,我们依据应用集成和研究热点的需求,对 PDS 和 SPASE 的属性交集进行遴选得到公共核心属性,如:观测平台、观测设备、观测区域、观测要素等,并在数据模型中构建统一的核心描述属性层。

(2) 分布式资源的全局发现

目前,要实现分布式的资源发现,通常有两种方式。第一种方式,是采用集中式的元数据查询,任何的资源请求都由该集中式的资源注册器解释和处理。系统通过与资源提供者进行周期性的全局同步实现资源全局发现。第二方式是,采用分布式的元数据查询,每一个资源提供者维护自己的数据资源和元数据。一个资源查询请求,被分发到每一个资源提供者。当系统收集所有资源提供者反馈的结果后,再合并并返回给用户。但是,这两种方式各有优缺点,在实现难易度上也不尽相同。

(3) 数据资源在分布式环境下的收割与同步问题

要达到虚拟观测的效果,就必须支持自动把新注册的资源信息实时同步到其它节点,使得每一个节点都能发现全局的数据资源。当前,关系数据库的实时同步技术,是一个业界尚未完全攻克的难点。

(4) 与空间模式的应用集成

空间模式之间 IO 标准差异大,物理数据存储格式不一,使得 VO 难以无缝地以数据仓库的形式为空间模式提供数据支持。为此,我们需要一种特殊的数据管道服务,不仅支持多协议的数据访问,还能依据管道两端的格式要求,自动实现格式转换。

(5) 面向科学问题的数据发现

常规数据检索,检索结果与学科数据需求的吻合度往往较低。因此,面向科学问题的增值型数据发现,将会是 VO 未来发展方向之一。空间科学 VO 将利用空间天气事件与观测数据的因果时序关联,构建具有学科内涵的数据发现。

(6) 空间科学元数据互操作问题

元数据的互操作就是在由不同组织制定与管理且技术规范不尽相同的元数据环境下,能够做到对用户保持一致性的服务,即列一个应用或用户来说,能够保证一个统一的数据界面,保证一致性与对用户透明^[39]。

元数据互操作性问题的产生主要有以下几个原因^[40]: ①不同的数据资源类型要求不同的元数据格式; ②不同层次和不同角度的应用会要求不同元数据格式; 不同领域甚至同一个领域内的不同机构,针对不同资源类型或为了不同层次或角度的应用,可能开发或应用不同的元数据格式及相应的元数据处理系统且已经积累了大量数据记录,由此导致多种不同元数据格式共存,而这些格式和标准互不

兼容,符合某种格式的元数据不能被其他格式所接受,从而导致不同数据库之间根本无法互相访问和检索。

然而,实现这一目标还存在很多的困难,这也是元数据互操作问题迄今为止还没有完全解决的主要原因。从宏观角度讲,元数据互操作涉及元数据各个结构层面的互操作^[41],包括:交换格式的互操作、标记格式的互操作、元素内容结构的互操作、元素语义的互操作、编码规则的互操作、数据内容的互操作,此外,互操作还要求通信层的支持。

(7) 分布式数据资源的快速加入和融合

丰富的数据资源是 VO 价值的载体和根本。因而,要提高 VO 的易用性,虚拟观测系统必须支持分布式数据资源的快速加入和融合。通常,如果采用集中式的资源注册方式时,VO 需要一个统一的数据模型,以支持统一的数据描述。这种方式的缺点是数据资源的加入和管理没有独立性。退一步而言,我们仍然可以考虑为数据资源提供统一的数据模型,资源提供者必须遵循这一数据模型,但是它们允许独立注册、管理和撤销数据资源。而最易于数据资源的快速融合的是,允许在一个 VO 里采用不同的数据模型,并且允许每一个数据资源提供者对自己的数据资源享有全部的权力。在该模式下,系统在进行资源检索发现时,必须为不同数据模型的 schema 进行解释和转换。

4.3 基本设计原则

为解决空间科学 VO 里的 7 个核心问题,我们针对这些问题,提出了对应的体系结构设计原则。我们认为,依循如下的几条设计原则,将会有助于设计出一个优秀的虚拟观测系统的体系结构。

首先,为实现空间科学资源与系统基础服务的高度整合,应采用 Service Oriented Architecture 基于服务的设计思想,即数据资源是服务 (data as a service),计算资源是服务 (computing as a service),应用资源 (application as a service)也是一种服务。

其次,系统应该具有网格节点和网格服务之间的关系应该是低耦合的。我们认为,节点和服务是自剥离的。节点作为一个或多个服务的载体,可装载或卸载服务;服务,作为虚拟观测台的功能单元,与节点并不存在必然的寄存关系。一

个节点可以部署一个或多个服务。一个服务可以部署在虚拟观测台中的任意节点。在虚拟观测台中，节点之间地位平等，集中式的主从关系被尽可能削弱。

第三，系统应该支持数据资源与 VO 快速融合的准入机制。要实现数据资源的快速融合和共享，必须分离传统的资源描述、资源存储和资源访问的三个基本要素。在空间科学 VO 中，资源提供者可以异地描述资源，异地为资源提供存储力，也可以异地提供资源访问接口。三者并不是不可分离的，它们可以不部署在同一个节点上。与此同时，系统还应该支持整合在线资源和离线资源，支持整合访问协议异构的分布式在线资源。

第四，无论采用分布式的资源注册器，还是集中式的资源注册器，空间科学 VO 都应该支持数据资源的全局收割和同步。在某个节点上，如果有新资源加入，系统应能够通过拓扑自组织服务，实现资源全局分布的更新和报告，达到快速同步的目的，以实现在任何一个节点上都可以够快速高效发现全局的所需的数据资源。

第五，面对来自不同学科的数据管理模型异构问题，空间科学 VO 应该致力于，通过在已有数据管理模型上增加抽象层，来实现数据模型的统一，从而降低数据资源提供者的负担，即允许数据资源提供者继续使用当前的元数据模型、传输机制和数据组织方式，不需要系统管理者创建全新的数据模型，也不需要资源提供者重新对数据资源进行整理和归档，更不要求在软件层面上，提供一个资源描述的中间转换机制。该原则的提出，旨在使系统设计保持最低的数据资源维护成本。

4.4 数据模型设计

为解决分布式数据资源的互操作问题，我们采用三层的公共数据模型（见图 4-1）：

（1）核心描述属性（见表 3），面向应用集成的需要来定义的，来自、但并不局限于标准描述属性，其目的在于支持快速的数据资源发现。

观测手段	天基	卫星平台 (fy2、double-star、ace、soho、themis、goes….)
		载荷类型

		(mag、sewpam、sis、epam、eit、lasco…)
	地	台站名称
	基	(海南综合观测站、怀柔太阳观测站、北京宇宙线观测站、十三陵地磁站 …)
		设备类型
		(磁通门磁力仪、中子堆、数字测高仪、激光雷达…)
观测区域	太阳、行星际、拉格朗日点、磁层顶、同步轨道、近地轨道、邻近空间….	
观测要素	太阳黑子、高能离子、高能电子、宇宙线强度、dst 指数…	

表 3 空间科学数据核心属性解释表

(2) 标准描述属性，采用支持国际数据资源的互操作的统一协议，包括行星数据系统（PDS）和空间物理归档、检索和抽取标准（SPASE）；

(3) 物理数据实体，用以支持底层物理数据访问和应用。

空间科学虚拟观测台的数据资源组织架构将以空间科学数据档案公共数据模型为基础进行构建，空间科学数据档案公共数据模型如图所示。

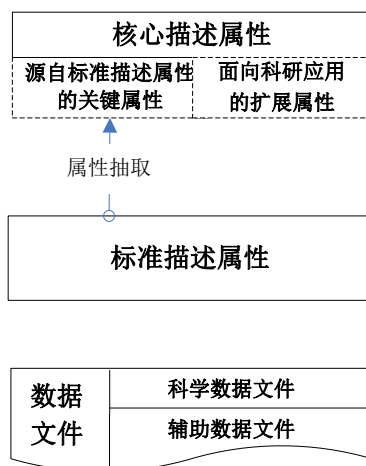


图 4-1 空间科学数据公共数据模型

公共数据模型是描述和组织空间科学数据资源（数据档案）的基本方法。公共数据模型阐述科学数据档案的数据元素组成，专业数据库和逻辑数据库构成方法（见图 4-2）。其中，专业数据库采用学科社区公认的描述属性（如观测目标、观测平台等）的属性值，所遴选的数据构建的数据集合。专业数据库一般是学科

数据的经典分类。逻辑数据库是面向典型应用需求和学科研究人员关注热点，采用同类核心描述属性（如关联事件、特定区域）属性值或属性值组合遴选的数据构成的数据集合。逻辑数据库可自由构建和动态生成。

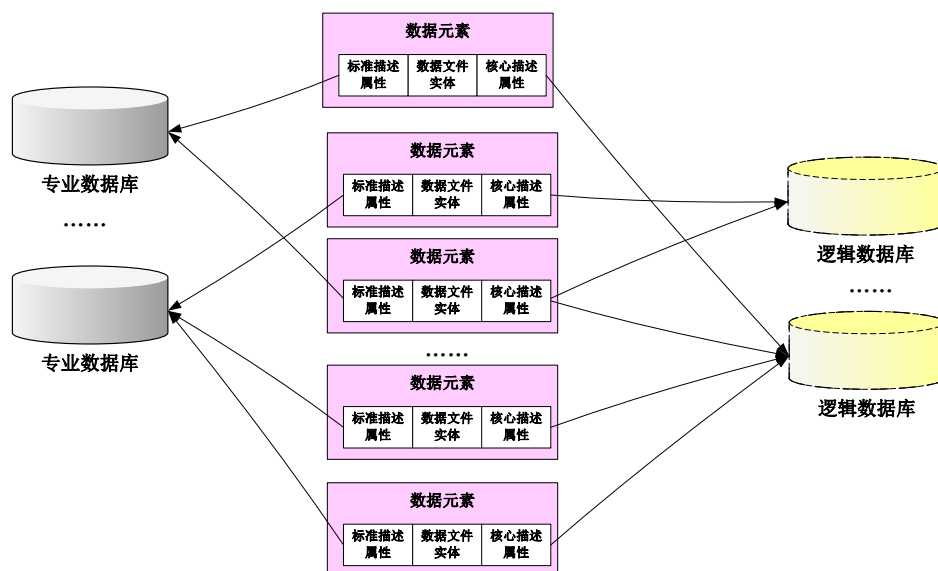


图 4-2 数据元素、专业库和逻辑库的关系

公共模型中最小的数据单元是数据元素，数据元素由数据文件实体和描述属性构成。描述属性由标准描述属性和核心描述属性构成。描述属性的值是元数据。标准描述属性是基于学科数据标准构建；而核心描述属性：面向数据网格应用集成的需要来定义的，来自于、但并不局限于标准描述属性。核心描述属性由平台类型、卫星型号、载荷类型、台站名称、设备类型、观测区域、观测要素、关联时间、时间等数据描述属性构成。

基于空间科学数据档案公共数据模型日地空间物理和行星科学数据资源组织架构分别如图 4-3 和图 4-4 所示。

日地空间物理数据资源的标准描述属性遵循空间物理档案搜索和抽取（SPASE）标准。按学科经典的分类方法，以标注描述属性中的观测区域为属性，遴选形成太阳活动、行星际空间环境、地球轨道环境、宇宙线观测、中高层大气观测、电离层观测、地磁观测、灾害性空间天气事件、卫星编目、空间环境效应共 11 个专业数据库。以核心描述属性中的关联事件属性构建 CME 事件关联数据库、重大磁暴事件关联数据库和质子事件关联数据库，以观测区域属性构建

临近空间数据库。其中卫星编目数据库、灾害性空间天气事件数据库的数据将作为核心描述属性的描述属性值。

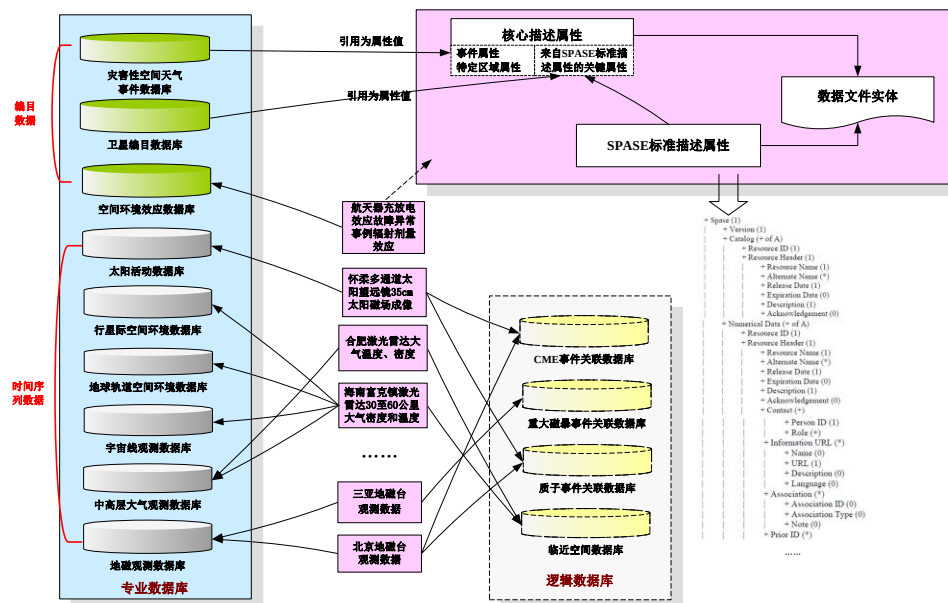


图 4-3 基于公共数据模型的日地空间物理数据资源组织架构

行星科学数据资源的标准描述属性遵循行星数据系统(PDS)数据标准。按标注描述属性中的观测区域为属性,形成火星探测专业数据库。

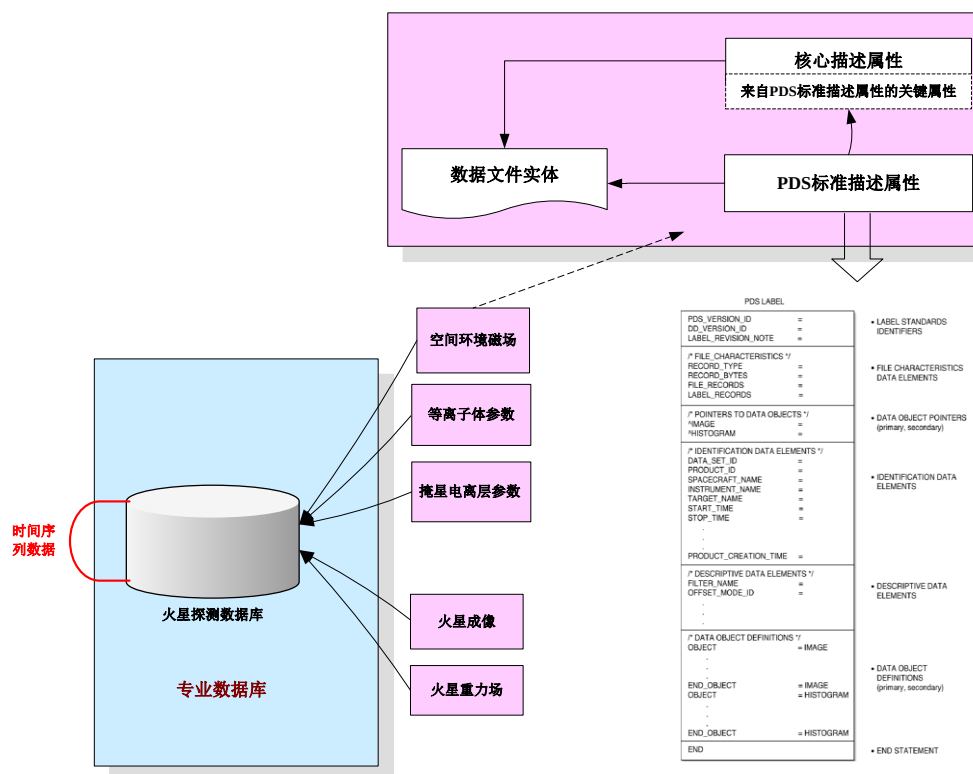


图 4-4 基于公共数据模型的行星科学数据资源组织架构

4.5 概念框架

空间科学 VO 概念框架（见图 4-5）的构成应包括资源、服务、应用和接入点等四类基本要素。

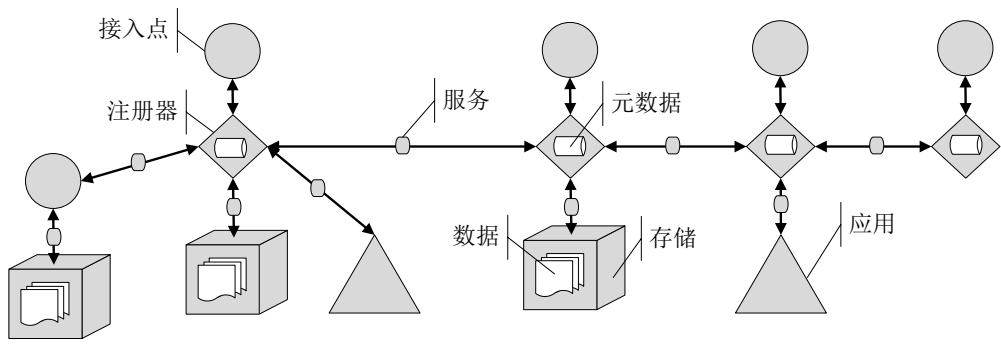


图 4-5 空间科学虚拟观测台的概念框架

资源是多样化的，数据、计算和存储等都属于资源；服务是 VO 的特定功能单元，它通过对外提供明确的接口来实现资源注册、发现、访问、同步等核心功能的封装；应用旨在为科研人员提供模式计算、结果分析和科学可视化等工具；接入点是 VO 和用户的桥梁，一方面用户通过接入点与 VO 进行交互，另一方面系统利用接入点进行资源信息收集，并将其存储在资源注册器中。

这四类要素通过自由组合与相互作用构成不同的功能实体，通过在 VO 物理节点上进行部署，形成对应的概念节点（见图 4-6）。

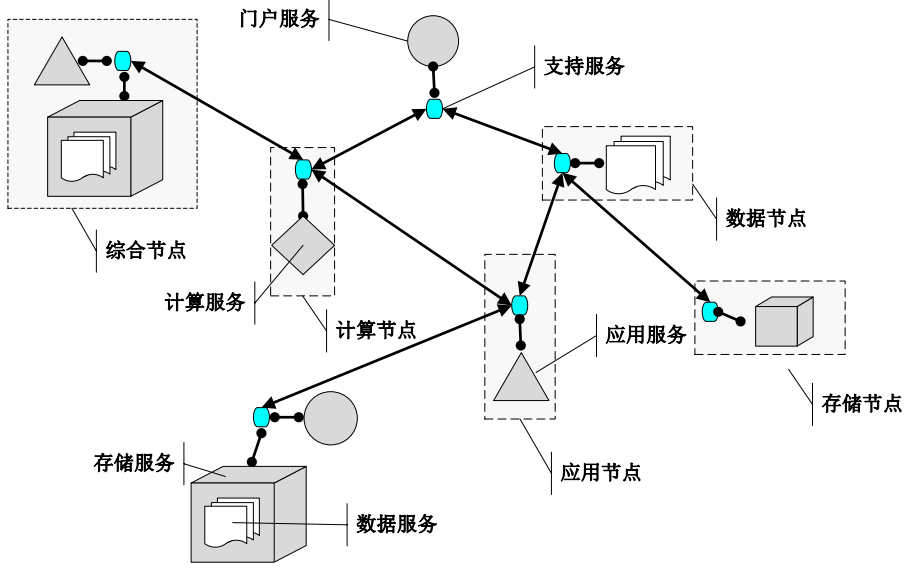


图 4-6 空间科学虚拟观测台节点与要素的耦合

在此原则下，一个物理节点可以同时包含资源、服务、应用和接入点等要素

构成综合节点；也可以只提供数据资源访问构成数据节点；或者以学科应用为唯一支撑构成应用节点，甚至仅依靠注册器构成注册节点。节点间的交互通过服务来实现，如：为实现资源的全局发现，资源注册器之间通过收割服务进行元数据的同步。

4.6 层次结构与协议

依据概念框架所阐述的原则，我们把空间科学 VO 的体系结构分为 3 层：资源层、服务层和应用层（见图 4-7）。

资源层是物理数据和元数据的汇集。“物理数据”层是分布式的文件数据系统，包括科学数据和辅助数据，用以支持物理数据访问和存储。“元数据”层又分为两层：标准描述属性层和核心描述属性层。其中，“标准描述属性”层是 XML 数据系统，针对行星科学和日地空间物理学分别采用 PDS4 和 SPASE 标准，以保留与 NASA VxO 的学科应用软件进行互操作的可能；“核心描述属性”层采用关系数据库技术构建，目的是实现资源的快速发现。“核心描述属性”主要来自于“标准描述属性”，但并不排除基于应用集成的需要，增加新属性的可能。

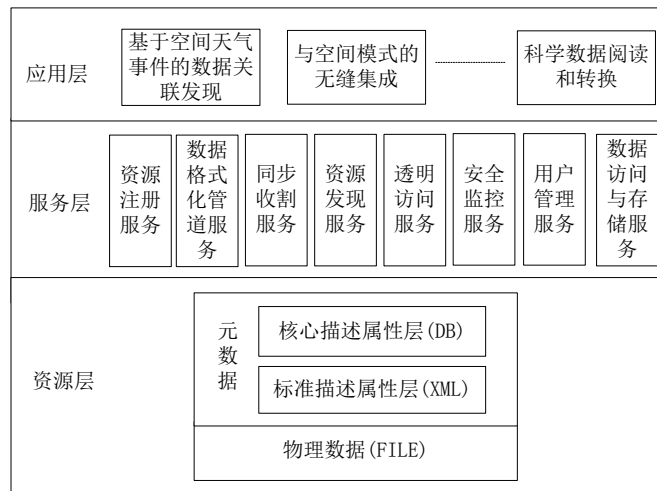


图 4-7 空间科学虚拟观测台的层次结构

服务层是面向服务的中间件集合，这些服务共同构成 VO 的基础设施。服务层主要提供 5 类服务：①面向资源，提供资源注册、发现、存储与访问服务。②面向拓扑组织，提供资源同步收割服务。③面向 VO 用户，提供用户管理服务。④面向系统安全，提供安全监控服务。⑤面向应用集成，提供数据格式化管道服

务。⑤面向虚拟化需求，提供透明访问服务。

应用层是一个学科应用池，提供数据计算、分析、可视化和知识发现的虚拟实验环境。它既支持单一学科应用，也支持多个学科应用间的耦合。应用耦合的 IO 标准异构问题，通过服务层的“数据格式化管道服务”解决。

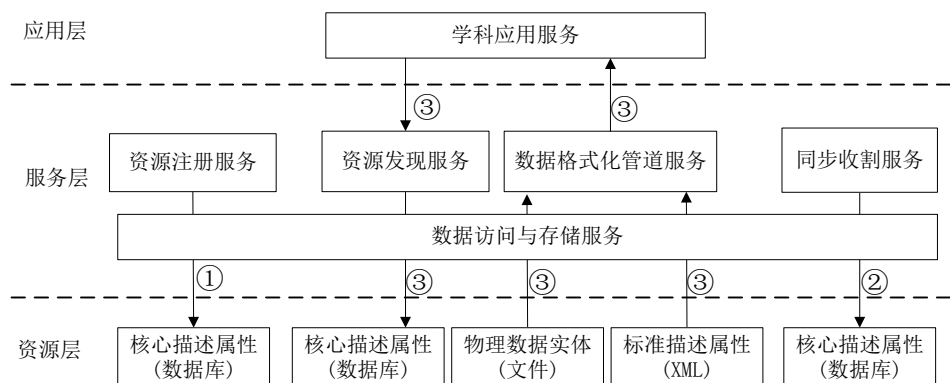


图 4-8 资源层、服务层和应用层之间的交互

资源层、服务层和应用层之间的交互，主要通过数据访问与存储服务提供文件、XML 和数据库等三个层次的读写接口实现。图 4-8 反映的交互关系如下：
①资源注册、审核、发布只需要对“核心描述属性”进行读写。②“同步收割服务”通过交换和合并各注册器的“核心描述属性”，最终实现资源全局同步。③资源发现服务检索通过“核心描述属性”来定位物理数据实体及其对应的标准描述属性，并利用“数据访问服务”将其传送到“数据格式化管道”。“数据格式化管道服务”依据“标准描述属性”的描述，对物理数据进行格式化后提交给学科应用服务，如 NASA SPICE 轨道分析应用就是在 PDS 标签的支持下，对物理数据进行访问、解析和抽取的。

4.7 服务模型

空间科学虚拟观测台服务是虚拟观测的基础设施，是构建应用服务的前提。我们认为空间科学虚拟观测台的服务模型（见图 4-10）需要包含如下服务：

（1）资源注册与发布服务。基于公共数据模型，实现数据资源的描述，审核和发布，使数据资源融入面向网格的虚拟观测台共享环境中。该服务的核心思想在于“分布注册、中心审核、分布发布”。

（2）资源发现服务。面向空间科学的研究需求和学科查询，实现快速高效

多维度的分布式数据资源检索和发现。

(3) 资源访问与存储服务。通过“存储服务”，为虚拟观测台提供资源、服务以及冗余备份的存储力；通过“访问服务”，支持在线资源和离线资源的快速共享，统一协议异构的用户访问，提高系统的兼容性和透明性。

(4) 数据高速传输服务。提供大数据量的高速传输，解决网格系统中资源和服务访问的负载均衡问题。

(5) 同步收割服务。同步收割服务的目标是实现虚拟观测台的“网格拓扑自组织”特性。该服务的核心思想是，充分利用和结合资源的注册、审核和发布过程，实现资源的全局同步。

(6) 网络安全和监控服务。网络安全和监控服务主要任务是负责空间科学虚拟观测台的安全。其服务功能核心主要体现在 3 个方面：第一，监视和记录网格系统的数据、用户、硬件、服务的情况，并进行统计、分析和挖掘；第二，基于监控日志，对已经执行的破坏行为自动进行操作溯源和状态回滚；第三，基于网格保护策略和安全访问协议，对将要执行的破坏行为进行规范和阻止。

(7) 冗余与透明切换服务。冗余和透明切换服务的目标是实现虚拟观测台的“服务自愈合”特性。主要包含两点：第一，通过拓扑自组织服务感知网格拓扑，并依据网格冗余策略，自动对物理数据、元数据和网格服务做冗余；第二，能够通过分析网络安全自保护服务的日志记录和监控结果，感知网格拓扑中节点和服务的健康情况变化，并自动基于网格冗余机制，进行服务访问透明智能切换。当任一节点的某网格服务出现问题时，系统能自动将该服务的访问负载切换到附近节点的同类服务上，提高网格服务的健壮性（见图 4-9）。

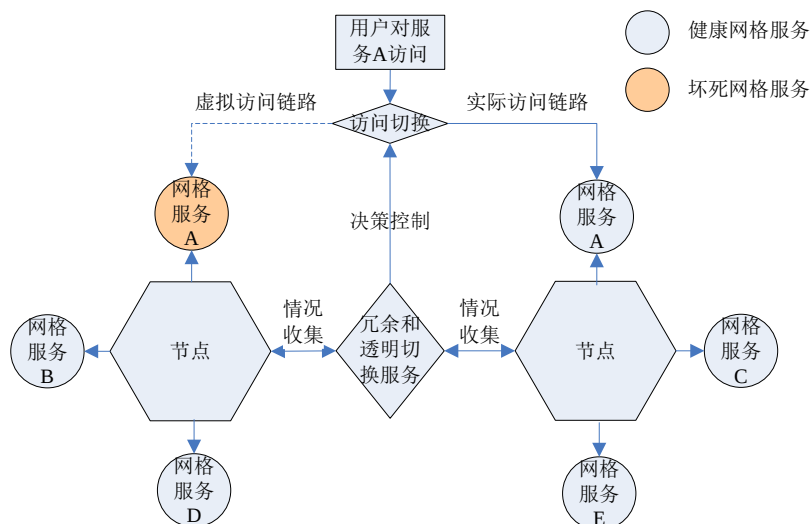


图 4-9 服务透明切换策略

- (8) 学科应用服务。面向学科应用，提供数据计算、分析、可视化、知识发现的应用工具和虚拟环境。
- (9) 网格用户管理。基于“统一注册，中心认证，局部授权”的三原则，解决用户注册、审核、授权和管理的问题，提高网格组织管理的灵活性。

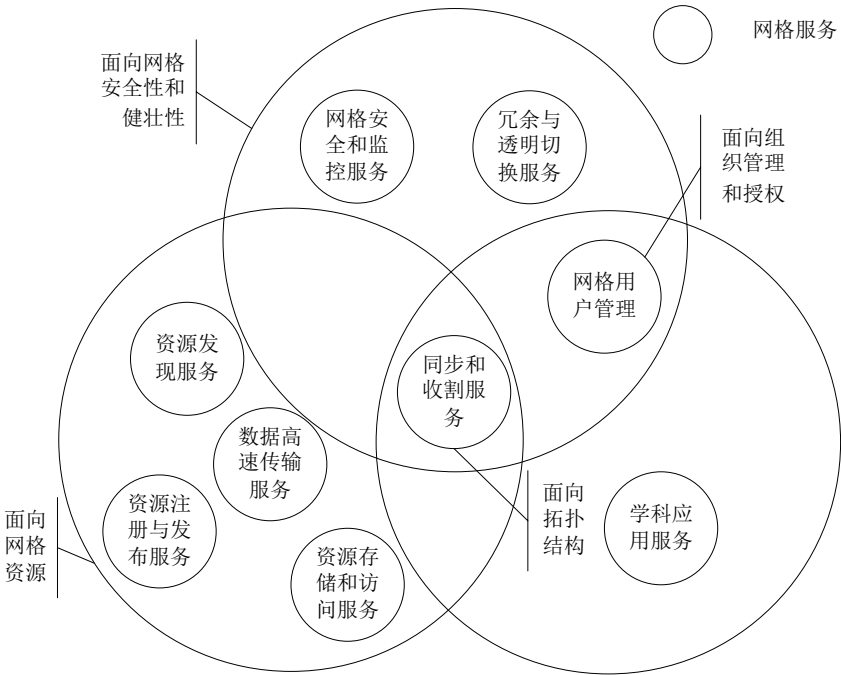


图 4-10 空间科学虚拟观测台服务模型

第五章 空间科学虚拟观测台的运行引擎设计与实现

5.1 概述

空间科学虚拟观测台的本质是融合虚拟观测理念的信息系统。如果能够基于上章中提出的体系结构概念模型，设计一个统一的运行引擎，以分解空间科学 VO 的业务逻辑、数据逻辑和交互逻辑，将会大大有利于减少开发时间，缩短维护周期，使得科研工作者能够把精力放在科学工作流和学科应用环境研究上。空间科学 VO 运行引擎的设计过程中，我们把虚拟观测系统的作用视为一个人机交互的过程。因而，虚拟观测系统的功能是通过空间科学数据应用环境的业务流与空间科学数据之间相互作用，实现对科学应用与研究（如：空间天气预测等应用）的支持。依据前文所定义的空间科学 VO 的概念框架和层次结构，我们把空间科学 VO 的运行引擎划分为三大部分：数据逻辑、业务逻辑和交互逻辑。这三个部分形成了上松下紧的“漏斗式”的空间科学 VO 运行引擎。

5.2 运行引擎的设计动机

在参与卫星数据信息系统建设的过程，研究人员和工程师发现传统的空间科学数据应用环境的建设和设计思路存在许多缺点。这些缺点产生的原因是，没有针对空间科学的数据特点，设计适合空间科学虚拟观测的系统框架和设计模式，因而，大量系统组件并不具有灵活性和重用性。

在数据模型上，我们以空间科学大气模式计算中的动态数据访问需求为例，美国海军实验室 MSISE Model 2001 大气模式计算，需要的参数之一是 F10.7 指数。F10.7 是表征太阳活动水平的参数，当太阳上发生射电爆发或者发生太阳耀斑伴随的射电爆发时，F10.7 会有较大幅度增加。作为一个固定性参数，MSISE Model 2001 计算要求静态的访问一个 F10.7 数据集。为此，我们只需要设计一个静态数据模型，以支持对 F10.7 的操作。但是，在 MSISE Model 2001 计算过程中，需要不断从不同载荷的观测数据读入对应时间的观测值。为解决该问题，我们可以有两种思路：①为每一个载荷观测数据集设计对应的静态数据模型；②设计一个动态数据模型，使得系统在运行过程中，能够动态改变该模型所指向的载

荷类型。

以基于观测设备、观测平台、观测要素、观测区域、观测时间、观测位置等核心属性的多属性关联数据发现为例,如果科研用户从观测设备检索到观测数据后,希望获得该观测数据的目标观测区域,那么传统的信息系统将从观测设备字典表,到物理观测数据表,再到观测区域字典表进行遍历。一个更好的做法是,把三个存储上物理分离的元数据集,按照一定的规则抽象成一个多维的数据模型,即:观测数据具有观测设备和目标观测区域两个属性,而每一个属性又有各自的属性。二级属性,即观测设备和观测区域的属性,虽然是间接和观测数据发生关联,但是在这个数据模型里,它们都被虚拟化为观测数据的一级属性,并且为该虚拟模型里的一级、二级甚至三级等属性都提供了直接操作的接口。

我们不仅希望在数据模型设计上获得可重用性,还希望在空间科学数据应用系统的科学工作流上实现重用性。以卫星数据处理为例,数据处理是卫星空间环境探测数据综合平台的核心功能,用于完成来自不同卫星型号种类繁多、格式各异、质量不一的大量探测数据进行有效、安全、可靠、规范的标准化处理,其核心业务是用来完成介入系统的不同卫星型号的原始数据文件进行文件排重、文件入库、文件拆包和时间码转换、数据记录排重、形成采样时间集合、轨道坐标计算、磁场坐标计算、探测原码到模拟量的转换、模拟量到物理量的转换和数据记录入库。目前需要处理的卫星型号已达到几十个,每一个型号卫星又挂载几个不同的探测仪器,不同的探测仪器探测不同的数据对象,每一个数据对象的处理方法也都不尽相同,不同数据对象在进行处理计算过程中可能要交叉使用才能完成各自的处理计算。因而,一个好的思路是,为每一颗卫星定制特定的工作流,用一定的语义规则表示。系统通过运行引擎解释该工作流定义文件后,按照顺序调入不同的数据预处理模块,用以解决动态工作流问题。

因而,我们希望设计空间科学虚拟观测台的运行引擎以达到五个目的:

(1) 在控制流程方面,作为系统应用骨架,空间科学 VO 运行引擎能够为确定虚拟观测系统的构件协作与交互衔接提供支持。基于引擎,系统能够构件任何的工作流。这些工作流类型应该包含:顺序型工作流、循环型工作流、条件型工作流,以及它们的组着。

(2) 在重用性方面,空间科学 VO 运行引擎致力于实现三种级别的重用:

内部重用，即在同一应用中能公共使用的抽象块；代码重用，即将通用模块组合成库或工具集，以便在多个应用和领域都能使用；应用框架的重用，即为专用领域提供通用的或现成的基础结构，以获得最高级别的重用性。利用空间科学 VO 运行引擎，能够在一个通用功能已经实现的基础上开始具体的系统开发。所以，运行引擎应提供了所有应用期望的默认行为的类集合。

(3) 在构建学科领域的专用框架方面，具体的学科应用和服务通过重写子类或者组装对象来支持应用专用行为。空间科学 VO 运行引擎除了强调的是软件的设计重用性和系统的可扩充性，将更侧重于面向学科领域（如专用学科数据模型）的重用。因而，它应该具有领域相关性，基础服务和学科应用根据运行引擎进行复合，构建出可运行系统。

(4) 利用空间科学 VO 的运行引擎，科研人员只需要关注 VO 中的科学应用功能，而不是科学计算、预测、分析、校验、可视化等流程如何在系统上运行起来。

(5) 基于空间科学 VO 的运行引擎，系统设计人员只需要关注海量卫星数据的抽取、清理、检索、映射、切割、旋转、嵌套、切片、钻取、坐标变换等操作，使之支持空间天气计算模式和可视化，而不是关注如何组织和存储卫星数据。

5.3 运行引擎的基本思想

空间科学 VO 运行引擎的基本思想是分离用户界面与基本应用程序的数据，将一个信息系统分成三个部分：①逻辑描述部分（model），也叫模型，用于描述该组件的逻辑状态和对这些状态信息的一些简单操作；②图形显示部分（view），也叫视图，即我们最终在屏幕上看到的表现形式；③控制流程部分（controller），也叫控制，用于描述逻辑描述部分如何对用户动作做出反应^[42]。

模型（Model）是软件处理问题逻辑在独立于外在显示内容和形式情况下内在抽象，封装了问题的核心数据、逻辑和功能的计算关系。它独立于具体的界面表达和 IO 操作。模型是应用程序的主体部分，表示应用程序的状态和业务逻辑，一个模型可以为多个视图提供数据。视图是用户看到并与之交互的界面，视图只是向用户显示相关的数据，并能接收用户的输入数据，但它不进行任何实际的业务处理。控制器接收用户的输入并调用相应的模型组件去处理请求，然后调用相

应的视图来显示模型返回的数据，与用户进行交互。

视图（View）是应用程序中用户界面相关的部分，是用户看到并与之交互的界面。通过访问模型对象中的数据，视图显示模型对象的状态。当模型对象中的状态发生改变时，视图对象所代表的用户界面内容也会相应地改变，达到内容与形式的一致；对于同一个模型对象，针对不同的用户请求，可以产生多个不同的视图对象。

控制（Controller）用于根据用户的输入，控制用户界面数据显示，更新模型对象状态。控制对象定义了应用程序的行为，负责视图对象与模型对象之间的同步，根据用户对视图对象的操作完成对模型对象的更新，并将模型对象状态的改变及时反映到视图对象上。

在这种设计理念下，系统的处理过程是，控制器接收用户的请求，并决定调用哪个模型来进行处理；模型根据用户请求进行相应的业务逻辑处理，并返回数据；控制器调用相应的视图来格式化模型返回的数据，通过视图呈现给用户。如果用户通过某个视图的控制器改变了模型的数据，所有其它依赖于这些数据的视图都应反映到这些变化。因此，无论何时发生了何种数据变化，控制器都会将变化通知所有的视图，导致显示的更新。

“视图-模式-控制”设计思路将数据表示、输入控制和数据处理分离开来，提高了软件结构的可扩展性和软件模块的可重用性。

5.3.1 面向计算的 Model 优化方案

但是，空间天气预报的大运算量和卫星海量数据存储的环境，传统部署和设计的两种方案会造成数据访问和计算的瓶颈。

一种方案是将 Model、View 和 Controller 都运行在客户端，由客户端处理所有的显示和业务逻辑，而服务器端仅作为普通的数据支持中心，负责数据读写。该方案的弊端是，客户端的计算量过大，客户端与服务器端之间的数据通信过于频繁，从而极大降低了系统效率。

另一种方案是将 Model 和 Controller 部署在服务器端，客户端只负责界面的显示，对于用户输入的解析和业务逻辑的处理都在服务器端进行。这种方式本质上类似单机式系统，没有充分利用分布在各地云端的计算能力。

为了合理地将模型中的计算任务调度到其他节点,我们根据模型中数据的不同特点,将模型数据分为动态和静态两部分,并据此将模型再分解为2层:面向动态数据的业务模型(Application Model)和面向静态数据的逻辑模型(Logic Model)。其中,Application Model 属于 Model 中面向用户的业务部分,而 Logic Model 则属于 Model 中面向底层的逻辑计算部分^[43]。

在该模式下,可以把 Application Model、Controller 和 View 部署在客户端,Logic Model 则部署在服务器端。客户端负责处理接近用户界面的业务流程,可以充分利用客户机的能力提供给用户丰富的交互体验;另一方面,服务器端负责连接数据库进行数据存取并处理复杂的逻辑计算^[43]。

5.3.2 面向存取的 Model 优化方案

传统的数据存取,往往采用动态的数据库查询,并解释返回的方式。但是,这种方式往往受制于数据库的存取效率,受制于表数据条目的量级,受制于机器的性能。随着计算机硬件的发展,计算机内存容量逐渐得以提高。完全动态的方式,忽略了结合硬件发展的态势,没有充分利用计算内存存取快速的优势。

为此,我们采用动态数据库访问和基于预存取的静态内存访问相结合的方式。

对于访问请求所涉及的数据量太大,以至于无法载入内存的情况,我们采用动态数据库访问方式,只有 VO 引擎接受到访问请求,才会向 DBMS 发起访问请求。当访问结束,VO 引擎立刻通过 DBMS 关闭该事务。

对于访问请求所涉及的数据量适合,可以载入内存的情况,我们优先把数据通过 DBMS 取出后,载入内存空间,再通过 Model 提供的操作接口,访问内存空间中的数据。但是,对于用户而言,它仍然感到仿佛正在动态访问数据库。

5.4 运行引擎的设计与实现

空间科学 VO 运行引擎主要由五部分构成: Protocol、Framework、Controller、View 和 Model。其中,Protocol、Controller、View 和 Model 围绕着 Framework,构成了运行引擎的核心(见图 5-1)。

Framework 基于 Protocol 解释用户的请求,按照解释结果载入对应的 Controller 和 View 到 Framework,并顺序运行。在运行过程中,Controller 和 View

通过 Framework 提供的 Model 引用函数，获得 Model 的句柄，实现对 Model 的各种操作，并最终把更新后的状态通过 View 返回给用户，实现和用户的交互。当用户依据更新后的状态发出新请求时，Framework 再次重复上述流程。

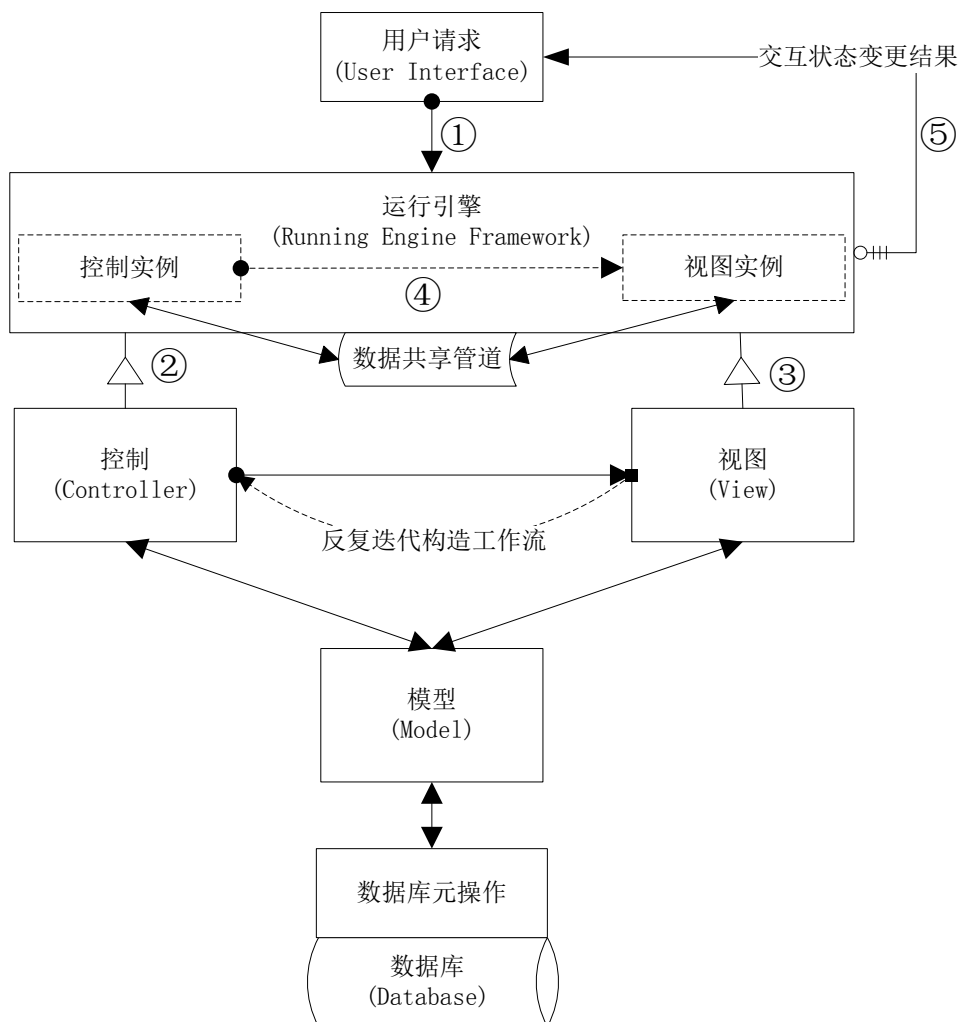


图 5-1 空间科学 VO 运行引擎设计方案

5.4.1 框架

空间科学运行引擎的框架(Framework)作为系统引擎运行的主体和骨架，在用户请求会话期间，首先把 Framework 实例化，加载到内存运行，并且在该会话结束前一直保留在内存中。Framework 主要支持如下功能：

- (1) 允许系统动态定义参数的传输协议；
- (2) 获取并分析用户的请求参数，明确要运行的动作 (action) 以及动作中的某一个方法 (method)，提取其他辅助参数；
- (3) 基于参数，定位要运行的控制模块 (Controller) 和视图模块 (View)；

(4) 依据请求, 动态载入 Controller 和 View 到 Framework 中, 分别实例化并按照从 Controller 到 View 的顺序执行对应的 method;

(5) 在常驻内存的 Framework 中, 开辟数据共享管道, 把 Framework 共享管道的引用交给 Controller 和 View, 使得 Controller 和 View 能够利用该管道定义、存储、共享和互访计算结果;

(6) Framework 提供动态调用 Model 的功能, 使得 Controller 和 View 在计算与输出的过程中, 能够获得所需数据的 Model, 使之具备对数据进行操作的权力;

(7) 运行引擎把每一次从 controller 到 view 的执行, 划分为 beforeInitiation, afterInitiation, beforeRender, afterRender, beforeOutput 和 afterOutput 等六个事件标志的流水线去运行;

(8) Framework 支持用户针对特定事件标志, 增加事件处理函数, 进行额外处理, 如拦截、检查、预加载等。

由此可见, Framework 的核心工作仅在于两点: ①正确载入控制模块和视图模块并按顺序执行; ②实现被载入模块之间的内存共享。

5.4.2 控制与视图及其完备性证明

运行引擎的控制(Controller)和视图(View)部分的核心思想是, 把虚拟观测系统视为一个人机交互的过程。每一个应用的业务流程, 都可以分解为一个或者若干个人机交互过程 (见图 5-2)。

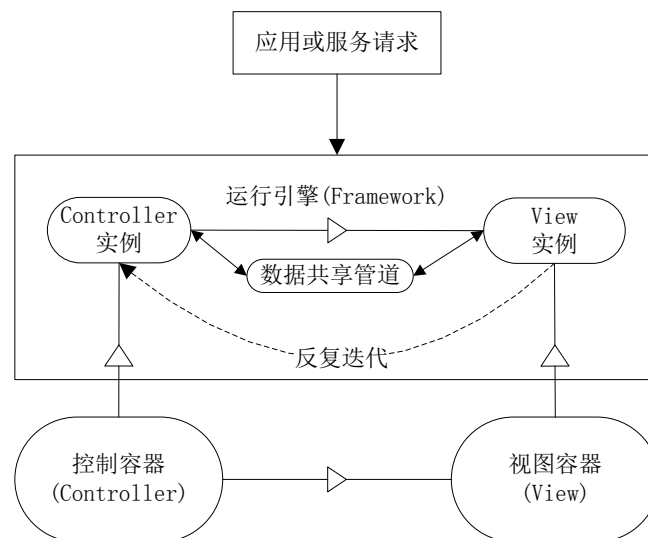


图 5-2 运行引擎的控制与视图运行时序

我们认为信息系统包含了控制（Controller）和视图（View）两种基本元素，和链接（Link）的一种运算。信息系统的业务流则可以定义为 Controller 和 View 在 Link 运算上的闭包：

$$\text{InfoSys} = \sum(C, V, \rightarrow),$$

其中 C 表示控制，即人机交互中的“机”；V 表示视图，即人机交互中的“人”； $C \rightarrow V$ 表示业务链接的前后顺承关系，即系统先执行控制模块，再把状态改变显示到视图； $\sum = \{\}^*$ 表示 C、V 在 $\rightarrow$ 上的闭包运算。

那么，人机交互中的 Controller 和 View 之间至少存在 4 中链接可能：

- Controller $\rightarrow$ View
- View $\rightarrow$ Controller
- Controller $\rightarrow$ Controller
- View $\rightarrow$ View

那么，信息系统可以重新表述为如下方式：

$$\begin{aligned}\text{InfoSys} &= \sum \langle C \rightarrow V | V \rightarrow C | C \rightarrow C | V \rightarrow V, \rightarrow \rangle \\ &= \{C \rightarrow V | V \rightarrow C | C \rightarrow C | V \rightarrow V, \rightarrow\}^*.\end{aligned}$$

我们对 InfoSys 进一步化简和推导，可以发现其除 Controller $\rightarrow$ View 之外，他三个链接模式都可以用 Controller $\rightarrow$ View 表示：

$$V \rightarrow C = \{C_1 \rightarrow V_1\} \rightarrow \{C_2 \rightarrow V_2\}, \text{ here } C_1 = \varepsilon, V_2 = \varepsilon$$

$$C \rightarrow C = \{C_1 \rightarrow V_1\} \rightarrow \{C_2 \rightarrow V_2\}, \text{ here } V_1 = \varepsilon, V_2 = \varepsilon$$

$$V \rightarrow V = \{C_1 \rightarrow V_1\} \rightarrow \{C_2 \rightarrow V_2\}, \text{ here } C_1 = \varepsilon, C_2 = \varepsilon$$

其中， ε 表示空模块，不执行任何功能。因而，InfoSys 可以最终化简为：

$$\begin{aligned}\text{InfoSys} &= \sum (C \rightarrow V, \rightarrow) \\ &= \{C \rightarrow V, \rightarrow\}^*.\end{aligned}$$

由上面的推导可知，空间科学 VO 的运行引擎只需要保持先加载系统的 Controller 再加载系统的 View 的方式，并通过这种 Controller $\rightarrow$ View 的叠加，就足以应付任何一种类型的工作流。

5.4.3 模型

模型(Model)的设计是反映一个信息系统运行引擎的学科特性的标志。如果是商业信息系统,其 Model 部分会针对买家、卖家、商品、评价、推荐等数据实体做针对性设计。

在空间科学 VO 运行引擎中,我们致力于使 Model 提供在海量卫星载荷数据上的抽取、清理、检索、映射、切割、旋转、嵌套、切片、钻取、坐标变换等操作,以支持 Controller 和 View 层面的计算和可视化对数据操作的需求,空间科学 VO 运行引擎的 Model,必须面向空间科学数据管理模型做针对性的设计。

空间科学数据管理模型分为两层:物理数据和元数据。其中,物理数据实体,主要包含物理观测数据实体和辅助数据实体。物理观测数据实体,主要指卫星载荷观测数据,反应观测对象的一个或若干物理要素在特定观测时间内的变化。辅助数据实体主要指对数据的说明,包括卫星、平台、仪器等参数的说明。元数据则主要用以支持底层物理数据访问和应用。元数据又分为核心描述属性和标准描述属性。核心描述属性的目的在于支持快速的数据资源发现,存储在在关系型数据库中;标准描述属性采用 XML 和 PDS LABEL 标签的存储方式,是文件型数据。

元数据和物理数据又可以按照档案组织方式,按照从粗粒度到细粒度,分为数据集粒度和数据文件粒度两种粒度。其中,数据集通常指在特定卫星上,某仪器上的一个载荷观测设备的历年观测数据集合,该载荷观测设备一般只观测一种物理要素;数据文件特指卫星数据产品,从产品级别上分,它可以是一级、二级、三级和四级卫星数据产品,从采集频率上分,也可以是 1 分钟快照、1 小时快照、1 天快照甚至 1 年的数据合成。

依据空间科学数据的管理模型,我们把 Model 划分为两大类:元数据模型和物理数据模型(见图 5-3)。其中,元数据模型按照从数据集到观测记录的粒度,划分为数据集元数据模型(静态数据集元数据模型、动态数据集元数据模型、可扩展属性的数据集元数据模型、可扩展观测记录的数据集元数据模型)和观测记录元数据模型(支持属性为单值的观测记录元数据模型、支持属性为多值的观测记录元数据模型、支持属性为树结构的观测记录元数据模型)。同样,物理数据模型也可以分为数据集模型和观测记录模型(物理观测数据实体模型、辅

助数据文件模型)。

在 Model 内部, 我们主要提供 4 种功能: (1) 为每一个具有不同数据内容和特点的模型, 设计合适的操作; (2) 实现高层操作的 SQL 语法自动翻译, 并确保语义的正确性; (3) 实现 Model 查询返回结果的数据值的类型自动转换; (4) 对查询结果进行判断, 如果数据量过大, 则采用动态从数据库提取方式, 这样可以避免内存占用率过高; 如果数据量小, 则采用内存驻留方式, 直到该 Model 被注销, 这样可以提高系统的数据读取速度。

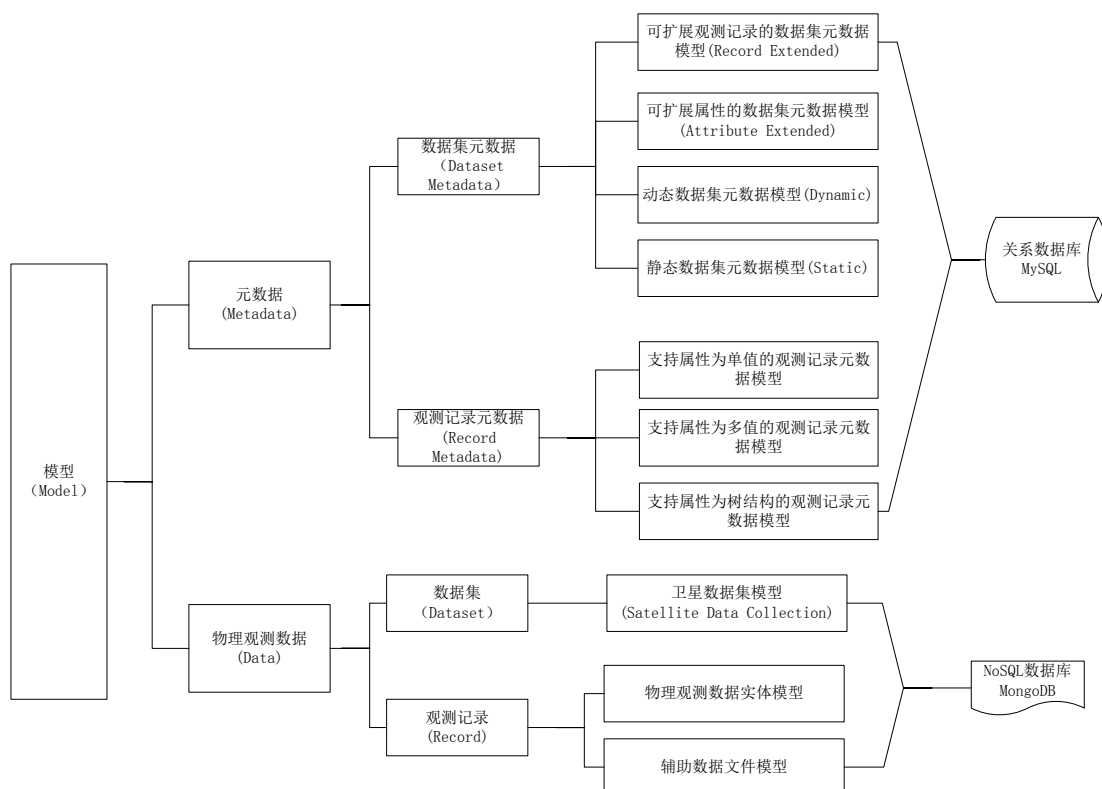


图 5-3 空间科学 VO 运行引擎的 Model 结构图

在 Model 外部, 我们致力于面向应用需求设计不同的虚拟 Model。每一个 Model 就好像一个具有操作的虚拟“仓库”, 仓库里存储着许多“商品”。这些操作可以对商品进行查找、统计、挖掘等。静态数据集元数据模型是最简单的模型, 用户通过重载该模型, 并指定具体的数据表以及表属性, 就能够通过它实现对数据的操作。动态数据集元数据模型, 是指在系统运行过程中, 允许用户动态重载和实例化一个数据集的元数据模型, 并动态指定该模型应该指向哪个数据表。可扩展属性的数据集元数据模型, 是指允许用户指定多个表的一个或多个属性通过 join 操作组成新的统一的模型, 用户就好像操作一个表一样, 操作多表属性构成

的虚拟表。可扩展观测记录的数据集元数据模型，是指当某些请求需要把多个具有相同描述属性的数据集合并起来，并联合检索的时候，用户可以利用它指定若干个表结构相同的表名，构成一个新的模型，这个新模型包含了多个表的所有数据条目。

通过这样的 Model 设计，科研用户可以面向学科应用需求，定制合适的 Model，实现虚拟直接操作，而不需要关注数据遍历和关联操作引起的后果。

第六章 空间科学虚拟观测台的原型设计和实现

6.1 概述

空间科学虚拟观测台的科学目标是,面向国家可持续发展、国防及重大科研应用对空间科学数据资源的应用需求,以中国科学院院内承担的重大空间科学项目、应用卫星空间环境探测和常规地面观测产生的数据资源为主体,广泛吸纳国际、国内相关重要数据资源,利用数据网格技术,屏蔽数据资源的存储形式、地理位置,分布式存储、管理、发布与共享空间科学数据资源,实现空间科学数据资源的注册、发现、监控、一站式访问、可视化应用与可动态接入的空间科学数据应用环境,为空间科学的研究工作者提供更为便捷的科研信息化环境(e-Science)。

空间科学虚拟观测台建设的提出,主要是针对空间科学研究信息化的需求。因而,它以科学数据库项目为依托,利用网格中间件技术,整合物理上分布的科学数据和资源,使得科研工作者可以方便、透明地访问和使用资源,为科研工作者提供高效、可靠和易用的数据应用平台,助力科研工作者加速科学研究进程,推动我国科研水平的提升致力于为空间科学研究人员提供易发现、易获取、易使用和易共享、高效率的 7*24 小时的集成应用环境。空间科学虚拟观测台还面向科学数据资源的拥有者,以合作共享为发展模式,科学规划,紧密结合需求,优选数据集,实现空间科学数据从点到面、从区域到全球、从零散到系统的集成与整合。

6.2 数据模型与数据库设计

6.2.1 数据资源内容体系

空间科学虚拟观测台集成的数据资源主要来自中国科学院院内承担的重大空间科学项目、应用卫星空间环境探测和常规地面观测活动,并以国外空间科学数据资源作为重要补充。数据资源按学科经典分类方法来分析,涉及日地空间物理、行星科学两大空间学科方向,其中日地空间物理方向涵盖太阳活动、行星际

空间环境、地球轨道环境、宇宙线观测、中高层大气观测、电离层观测、地磁观测、灾害性空间天气事件、卫星编目、空间环境效应等内容。

所涵盖的空间科学数据资源具有以下特点：

(1) 数据的空间、时间尺度变化范围大。从地面 2-30km 以上直至太阳，它由物理性质和结构不同的太阳大气、行星际介质、地球的磁层、电离层、中高层大气 5 个空间区域；时间尺度从几分钟到 11 年太阳活动周变化。

(2) 数据的对象和参数多。包括太阳活动、日冕物质抛射和太阳质子事件、行星际太阳风扰动、行星际磁层方向变化、磁层亚暴和磁暴、粒子沉降、辐射带变化、电离层扰动和闪烁、中高层大气密度和温度、风速和成分等。

(3) 数据的时序特征强。天基和地基观测相配合，多手段、多点位观测相结合，对太阳、磁层、电离层、地磁场进行有效、连续的监测是获取连续可靠空间天气数据的主要手段，某一时刻的监测数据是瞬时的空间天气变化记录，不可复现。

(4) 数据的相关性强。日冕物质抛射、行星际风暴、地磁暴、银河宇宙线暴等空间天气事件，空间环境效应、空间目标监测等存在很强的时序因果关系。

(5) 以空间科学数据的时间特性作为分析基准，空间科学数据可以概括为两类数据：①时间序列数据，即对一个对象长时间的连续观测记录；②编目数据，即针对某一个特定事件的记录。

6.2.2 数据资源组织架构

空间科学虚拟观测台的数据资源（见图 6-1）主要分布在中科院空间中心、国家天文台、中科院地质地球所和中国科技大学。

中科院空间中心建设有地磁、电离层、宇宙线和中高层大气等空间天气要素常规地面观测系统以及神舟飞船探测、风云系列卫星探测、资源系列卫星探测、实践系列卫星、双星卫星等空间环境天基探测系统，可获得地面磁场强度、电离层特征参量、电离层 TEC、大气风场、大气密度和温度、气辉、大气分子密度、气溶胶消光系数、宇宙线中子计数、高能粒子辐射、轨道大气环境等一系列天地基数据资源；正在开展的中俄火星探测计划获取的行星科学探测数据资源可作为空间科学主题数据库的重要数据来源。空间中心挂靠管理的世界数据中心（WDC）

中国空间科学学科中心和欧洲 CLUSTER 数据中心中国分中心是代表我国进行国际间空间环境数据交换的窗口，长期开展国际数据资源的引进和交换工作，积累了大量的国际空间科学研究领域具有权威性的数据资源，如国际著名的大熊湖天文台太阳观测数据、磁场观测数据、地磁指数数据、灾害性空间天气事件数据、卫星编目数据和 GOES、ACE、SOHO、WIND 卫星的太阳风等离子体、行星际磁场、高能粒子、高能粒子辐射行星际轨道、高空磁场数据等，这些国际交换与镜像数据可为空间科学主题库数据资源的重要补充。

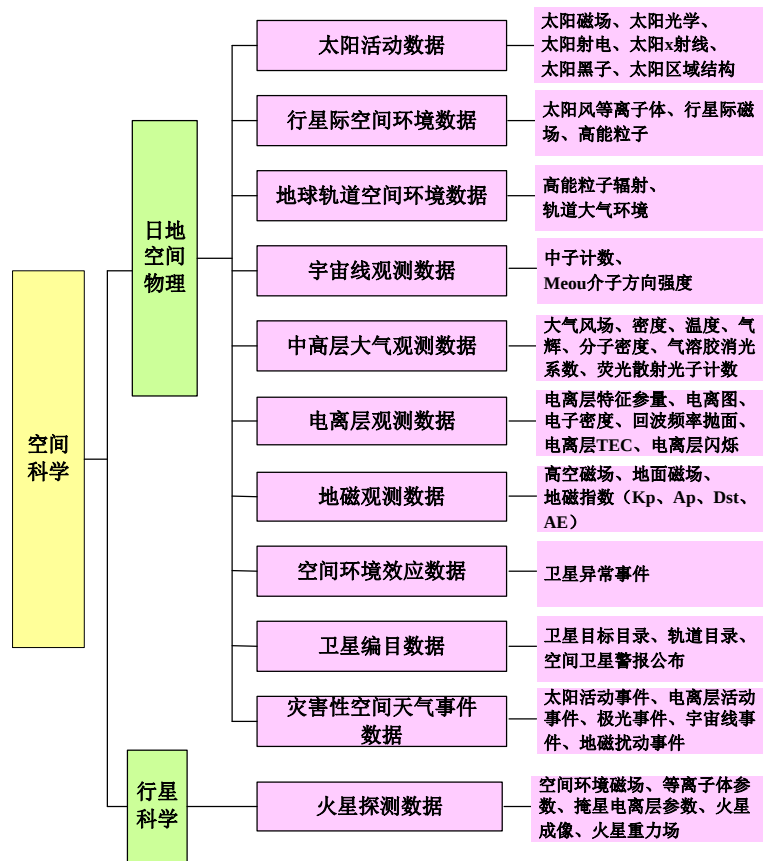


图 6-1 数据资源结构图

中科院国家天文台、中国科技大学和中科院地质与地球物理所作为空间科学主题数据库的建设参加单位，提供的数据资源主要以其单位常规地面观测台站所产生的数据资源为主（见图 6-2）。其中，中科院国家天文台在北京怀柔建设有太阳光学和太阳磁场地基观测系统，可获取太阳光学观测、太阳磁场观测数据；中国科技大学在合肥建设有激光雷达观测台站，可获取对流层-平流层大气气溶胶消光系数、平流层-中间层大气密度和温度、中间层-低热层内金属层 Na 原子的数密度等中高层大气数据。中科院地质与地球物理所在漠河、三亚、北京建设

有地磁观测系统，可获取地面磁场总强度、地磁三分量数据。

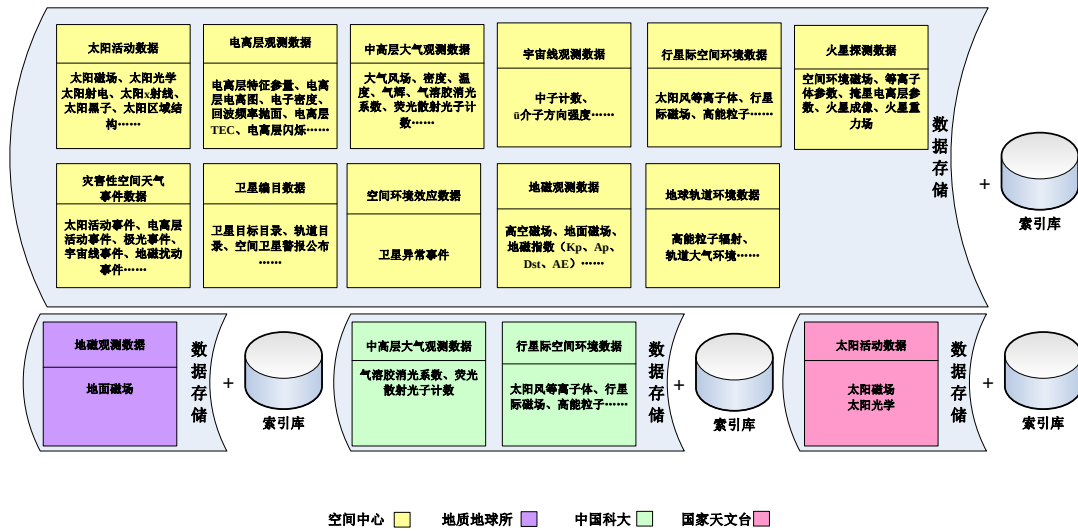


图 6-2 数据资源存储方式

6.2.3 数据库设计

空间科学 VO 系统成功运用的深度，依赖于有效的元数据模型和数据库结构。元数据为 VO 的数据充当路标，从而能够对海量物理观测数据进行高效管理、控制变更和分发。全面的元数据管理保证了空间科学 VO 具有高质量的信息，并提供充分的扩展性，能满足新信息需求和数据源增加。元数据实施是信息集成中的一部分，最重要的工作是将系统中用户、业务、技术、数据等的元数据进行整合利用。

	业务元数据	角色元数据	观测记录元数据	技术元数据
顶层	核心服务	用户分类	数据集	系统
底层	服务实体与业务功能	用户属性	数据记录	技术对象与技术元素

表 4 空间科学 VO 数据库设计模型与元数据划分

空间科学 VO 的元数据模型主要分为四部分（见表 4）：业务元数据，角色元数据、技术元数据和物理观测记录的元数据。

从业务角度分析，空间科学 VO 的基础设施框架包含如下核心服务：门户服务、资源注册；资源发现；网格用户管理；网格监控和统计服务；单点登录服务。每类服务都具有其元数据，如：资源注册服务就应该记录资源提供者、数据录入

时间、申请审核时间、审核管理管理员、审核时间、审核结果等属性。

从用户角色角度分析，在空间科学 VO 中有 5 类角色：科研工作者；资源提供者；审核管理员；发布管理员；用户管理员；每一个用户，可以同时拥有以上一个或者多个角色。在用户登录的时候，系统将根据登录用户的权限为用户创建相应的角色的实例，并将他保存在会话中，当用户以后进行操作的时候，系统将根据用户的请求获得操作对应的角色的实例，然后通过角色的实例来进行操作。因而每一个用户也应该具有元数据，如：姓名、单位、访问权限等。

从观测记录角度分析，空间科学 VO 的数据组织按照粗粒度到细粒度分为数据集和观测记录。数据集包含观测要素、卫星平台、观测设备、载荷类型、观测对象、单位标量等属性；观测记录包含：观测时间、观测位置等属性。每一个属性又对应一个标准化的数据字典。技术元数据，则侧重于记录数据访问协议，访问地址组织方式，服务的参数传递方式、参数名称以及参数意义进行记录，以保证系统的正常运维。

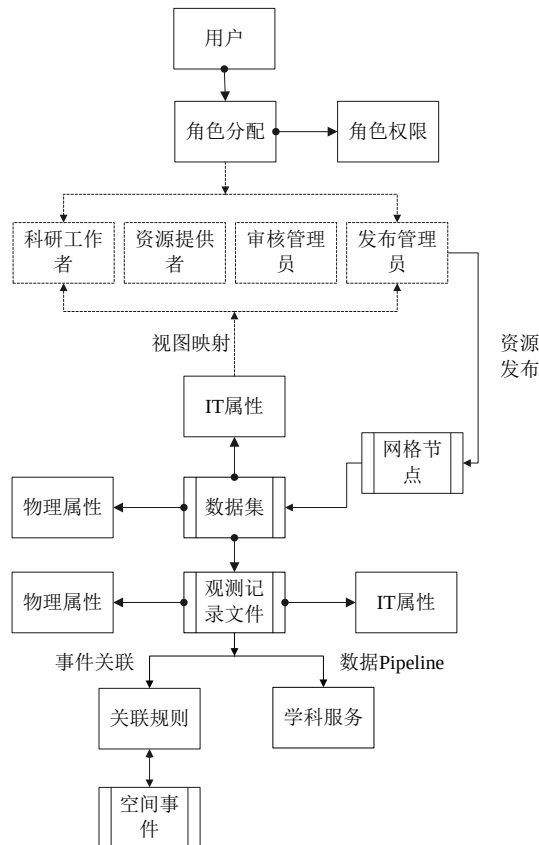


图 6-3 空间科学虚拟观测台数据库逻辑设计

		数据检索和访问权限的限制
		服务访问和使用的权限限制
		数据备份
	一站式单点登录	票据产生
		票据分发
		票据验证
		票据销毁
	监控服务	用户监控
		数据监控
		服务监控
		硬件监控
面向科研工作者	资源注册服务	数据资源录入
		数据资源审核
		数据资源发布
	资源发现服务	基于核心属性的数据发现
		基于目录导航的数据发现
		基于关键词的数据发现
	资源存储与访问服务	资源存储
		资源访问
	学科应用服务	科学数据可视化
		空间事件数据挖掘
		空间天气模式计算
		学科数据在线转换和阅读
	门户服务	网格服务交互接口
		信息发布
		用户个性化定制
		用户注册管理

表 5 空间科学虚拟观测台核心服务分析

6.3.1 门户服务设计

门户服务是面向空间科学虚拟观测台向科研工作者和系统构建者设立的,它为两类用户提供信息交互和功能展示的平台。通过门户服务,系统向用户提供用户登录验证、资源注册、资源发现、资源访问、资源下载、学科应用、个性化定制等服务的访问接口;通过门户服务,系统把在物理上分布的数据网格服务资源和数据资源都集成到统一的 Portal 中,屏蔽网格系统的分布性,使得服务和数据的获取集成化和透明化。

门户服务主要分为 4 个子服务,分别为:网格服务交互接口、信息发布、用户个性化定制和用户注册管理。

(1) 网格服务交互接口,指为实现分布在不同节点的网格服务在 Portal 上的集成,系统通过网格交互接口架设某一网格服务从 Portal 到服务部署点的访问通道。网格服务交互接口主要包含 5 个部分:

- 用户访问认证服务交互接口:指用户登录网格系统,通过提交用户名、密码等用户信息到验证系统进行身份辨别。
- 资源注册服务交互接口:指资源提供者,利用注册服务交互接口,调用资源注册器,将数据资源描述信息,注册到空间科学虚拟观测台。
- 资源发现服务交互接口:指门户上提供资源发现服务的 2 种发现模式,即基于目录树分类片面的数据资源发现和基于检索的数据资源发现。用户通过资源发现服务交互接口,发现服务发送请求,最后由发现服务截获命令并解析执行。
- 资源存储和访问服务交互接口:指用户进行资源发现后,通过资源存储和访问服务交互接口,向数据资源存储点发送请求,架构在数据资源存储点上的资源存储和访问服务解析请求,把数据资源返回给用户。
- 学科应用服务交互接口:指用户通过学科应用服务交互接口,进行科学数据计算、分析、统计和可视化。

(2) 信息发布,指系统管理员或者系统,通过信息发布子服务,把数据资源的更新情况、政策及标准规范、科普推广等用于维护系统正常运作的信息发布到 Portal 上,并依照需求执行信息的增、删、改、查等管理任务。信息发布子服务主要包含两个功能模块:信息发布和信息管理。

(3) 用户个性化定制, 指门户允许用户向系统提出个性应用需求, 系统依据不同用户特点, 个性化定制和突出用户所关注的主流数据和服务。用户个性化定制主要分为 2 个功能模块: 数据篮子和个性化发现。

- 数据篮子: 指用户通过数据篮子锁定常用的重点关注的数据, 以使得用户能够快速获得重点关注数据资源。
- 个性化发现: 指门户允许用户依据常用检索需求, 基于底层检索原语, 定制个性化检索。通过将自己常用的集成检索范围定制成 Panel, 节省在检索时重组检索逻辑的时间消耗。

(4) 用户注册管理, 指门户提供向科研工作者和系统管理者提供注册管理接口。通过它, 科研工作者可按照网格系统规范和准入机制, 提交必要信息申请成为数据网格的用户; 通过它, 系统管理员能够对申请用户进行审核、管理和授权。用户注册管理分为 3 个功能模块: 用户注册、用户审核和管理、用户授权。

6.3.2 资源注册服务设计

资源注册服务是空间科学虚拟观测台向数据资源提供者的一种服务。为数据网格提供数据资源的提供者, 可以依据空间科学虚拟观测台制定的公共数据模型和描述标准, 对数据资源进行描述。数据资源经过资源注册服务的录入、审核和发布三阶段, 最终融入开放的网格共享环境。

为更好的反映描述信息的数据流和数据状态变化, 我们设计 3 种不同的状态对资源注册过程中, 数据状态的变化进行区分 (见图 6-5)。它们分别是: ①等待审核状态, 指数据资源对应的核心描述信息刚注册到数据网格, 被系统缓存在“元数据缓冲器”。在该状态下的核心描述信息, 是未经审核的, 非正式的, 不会被用于数据发现服务。②审核通过状态, 指经过系统管理员审核, 判定为合格的, 存储在“全局元数据容器”, 并被推送到各子节点的“元数据容器”保存的核心描述信息。在该状态下的核心描述信息, 是经过审核的, 正式的, 将会用于数据发现服务。③审核不通过状态, 指经过系统管理员审核, 判定为不合格, 无法通过的核心描述信息。该状态下的核心描述信息, 虽然经过审核, 仍然为非正式的, 不会被用户数据发现服务。

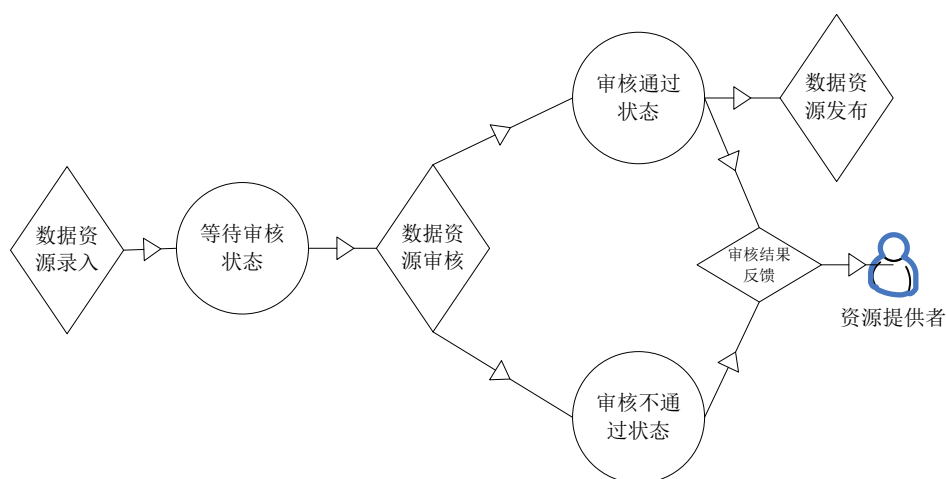


图 6-5 资源注册阶段的数据描述信息状态转换图

空间科学 VO 把资源注册服务分为数据资源录入、数据资源审核和数据资源发布等三个阶段（见图 6-6）。其中，数据资源录入，是指依据数据资源的运行类别不同，采取不同的录入策略，实现资源从离线到在线的转变。数据资源审核，主要完成审核任务生成、元数据审核、审核结果反馈、元数据状态转换等功能。数据资源发布子服务，则包括数据访问权限授予、数据资源发布地选择、同步元数据的审核结果等工作。

（1）数据资源录入，指数据资源提供者通过资源注册器，把资源描述信息录入空间科学虚拟观测台子节点的元数据缓冲器，系统检测到新数据的注册后，通过触发机制，启动拉模式收割器，把各子节点的元数据缓冲器中的属于等待审核状态的描述信息收割到全局元数据容器。

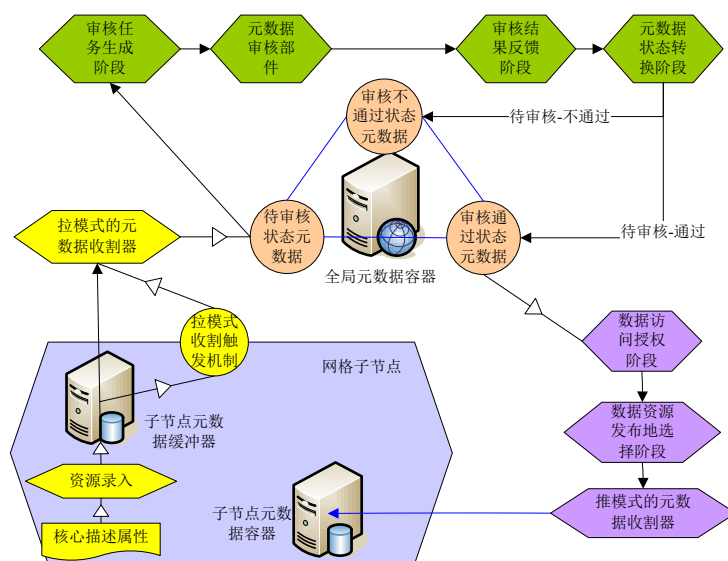


图 6-6 数据资源注册服务流程设计

依据数据资源提供者的不同,我们分为两种情况:数据资源在线(见图 6-7)和数据资源离线(见图 6-8)。其中,对于在线数据资源,如来自中科大、空间中心、国家天文台、地质地球所等 4 家共建单位的数据,数据资源提供者已经为“物理数据实体”和“标准描述属性”提供独立的存储力,物理数据资源已在线,资源提供者能独力确定访问地址,并且表达达到描述信息中;数据资源的提供者,只需要录入“核心描述属性”即可。对于离线数据资源,数据资源提供者不但需要录入数据描述信息,还要求空间科学虚拟观测台为其提供存储空间,以存储数据资源,由空间科学虚拟观测台来实现数据资源的从“离线”到“在线”的转变。

因而,在数据资源录入阶段,系统的主要目标是录入数据资源的描述信息。为此实现该目标,我们设计了三个核心部件:元数据缓冲器、拉模式收割器、全局元数据容器。元数据缓冲器设立的目的有 2 点。第一,“缓冲”清晰表明,元数据缓冲器存储的是新注册的数据描述信息,其状态为等待审核状态,属于未经审核的非正式的元数据;第二,元数据缓冲器为系统提供自动检测是否有新数据资源注册的访问接口,当发现有新的数据资源注册进来,则触发拉模式收割器对待审核元数据容器进行收割和同步。拉模式收割器设立的目的是使系统能够依据新数据资源注册的触发,自动收割各子节点的新注册的数据资源描述信息,统一存储到全局元数据容器,为数据资源的统一审核打下基础。全局元数据容器是实现全局数据资源同步的核心部件,在全局元数据容器的描述信息分为 3 种状态:待审核、审核通过、审核不通过。它在不同的阶段提供不同的功能。在“数据资源描述信息录入”,其功能是:实现对网格系统各节点数据资源描述信息的全局收割和集中,为系统管理员对数据资源进行统一审核,提供底层数据操作接口,从而转入数据资源审核阶段。

(2) 数据资源审核(见图 6-9),指系统管理员,针对数据资源的描述格式规范性、数据资源的可访问性、数据资源的描述内容和对应物理数据的吻合度、数据资源访问权限设定等影响数据资源质量的核心因素分别对数据资源进行审核评分,对于审核不合格的,驳回发布请求,把该部分数据资源描述信息的审核状态标识为“审核不通过”;对于审核合格的,把该部分数据资源描述信息的审核状态标识为“审核通过”,转入数据资源发布阶段。

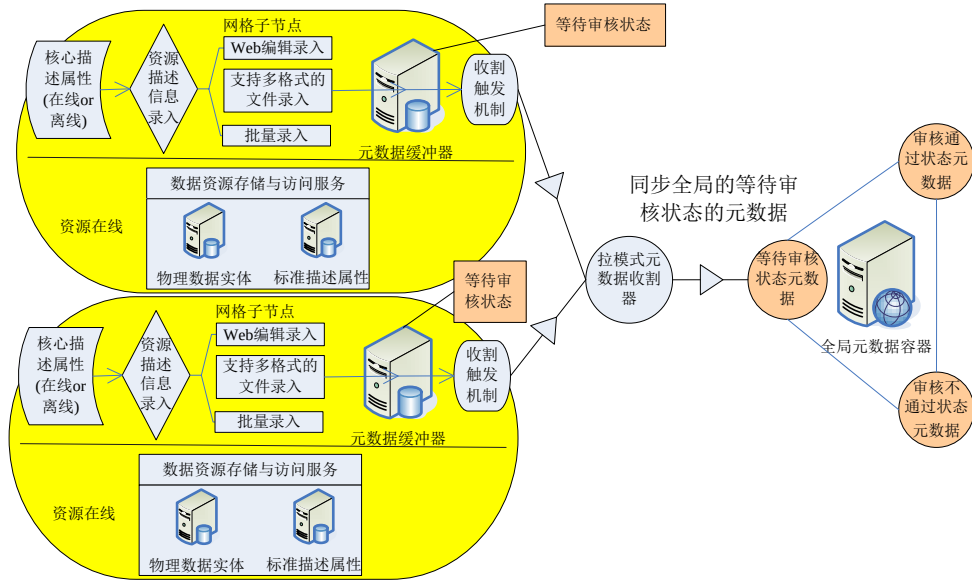


图 6-7 在线数据的资源录入

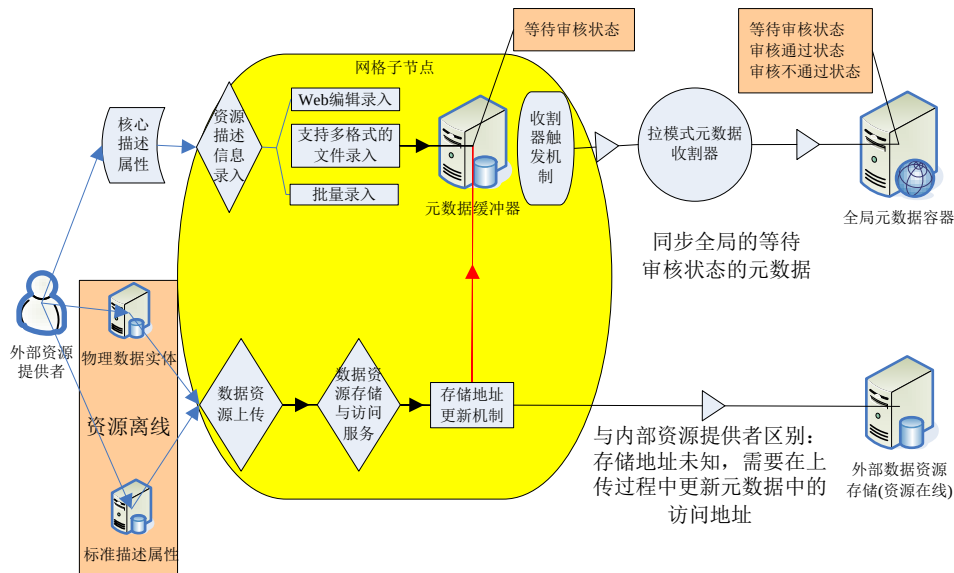


图 6-8 离线数据的资源录入

为清晰表达数据资源审核的过程，我们把数据资源审核划分为 4 个阶段：① 审核任务生成阶段，指系统通过新审核任务触发机制，检查“全局元数据容器”中是否有新的“等待审核状态的元数据”到来，如果检测结果是有，则针对该部分“元数据”，触发进入审核任务生成阶段，为系统管理员生成一个资源审核任务。②元数据审核阶段，是指系统管理员进入审核元数据的门户，分别执行如下工作：元数据描述的格式检查、查看数据文件实体的链接地址、下载数据文件实体、检查数据实体的元数据和对应的数据实体的吻合度、更改数据实体的访问检

索权限等。；③审核结果反馈阶段，审核结果反馈阶段，是指经过元数据审核阶段，系统管理员把审核结果和原因反馈给数据资源提供者。④元数据状态转换阶段，是指把审核结果反馈到“元数据状态”，把元数据设置为“审核通过”或者“审核不通过”。

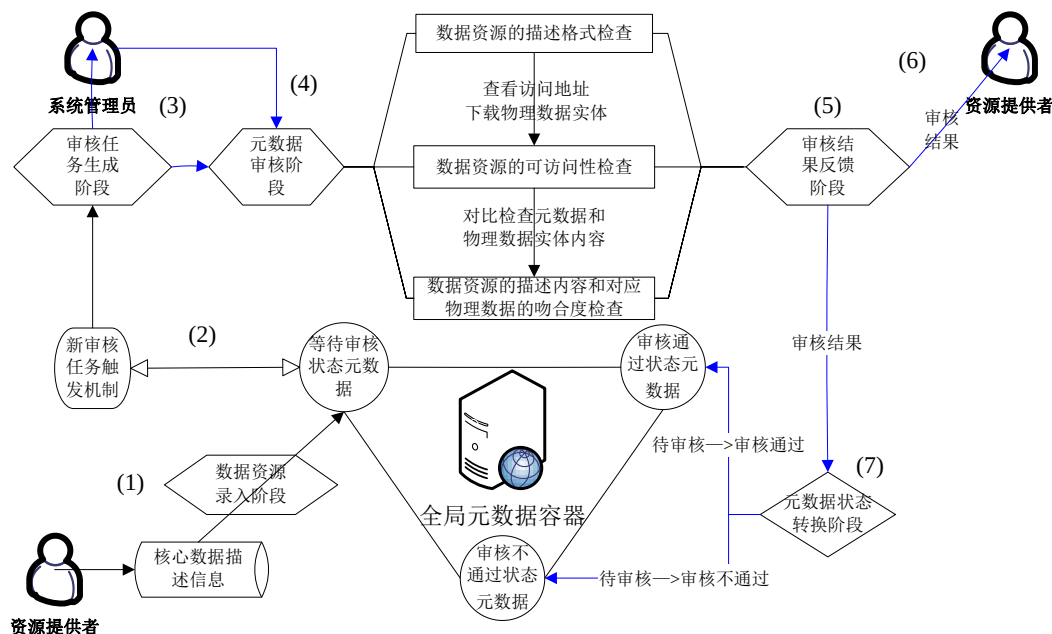


图 6-9 数据资源审核

(3) 数据资源发布, 指对于审核合格的数据资源, 系统通过元数据状态转换阶段, 把该数据资源在全局描述信息容器中对应的描述信息的审核状态转为“审核通过”, 系统管理员为该部分数据资源, 设定数据检索访问权限, 并依据需求选择数据资源的“发布地”(即子节点), 通过“推模式收割器”, 把该部分数据资源的描述信息推送到对应子节点元数据容器中。子节点的元数据容器, 装的是的审核通过状态的元数据, 用以支持局部资源发现; 全局元数据容器, 用以支持全局资源发现。

依据上述定义，数据资源发布（见图 6- 10）可以划分为 3 个阶段：数据访问权限授予阶段、数据资源发布地选择阶段和同步元数据阶段。

在数据访问权限授予阶段,考虑到数据网格数据资源的最终目的在于资源共享,从统计和概率上看,大多数用户拥有资源的检索和访问权限。因而,我们对于特殊资源,采用黑名单制度,来限制用户对资源的访问。这样的黑名单,是属于稀疏数据,存储空间小,限制访问的实现复杂性也将会大大降低。

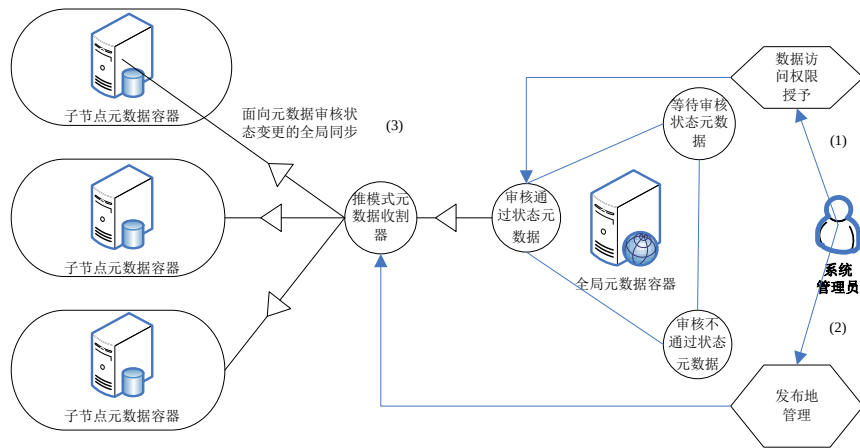


图 6-10 数据资源发布

数据资源发布地选择阶段，是指系统管理员依据需求，选择全部或者部分子节点，作为数据资源发布地，把审核通过的元数据推送回到选定的子节点元数据容器。

同步元数据阶段，是指推模式元数据收割器，依据系统管理员所选择“数据资源发布地”，把“审核通过”的元数据推送到对应的子节点。因而，“拉模式的收割器”和“推模式的收割器”其核心任务是有区别的。前者目的是实现新注册数据资源的描述信息的全局同步，后者目的是实现描述信息的审核状态变更的全局同步。

6.3.2 资源发现服务设计

资源发现服务（见图 6-11）是空间科学虚拟观测台门户系统为用户提供的核心功能之一，其目标是使得用户能够方便地发现和获取自己所需要的分布在空间科学虚拟观测台中的数据资源。空间科学虚拟观测台的资源发现服务将会紧密结合空间科学数据的学科需求，针对空间科学数据的物理要素内在特征，实现多维度的分布式数据资源发现。

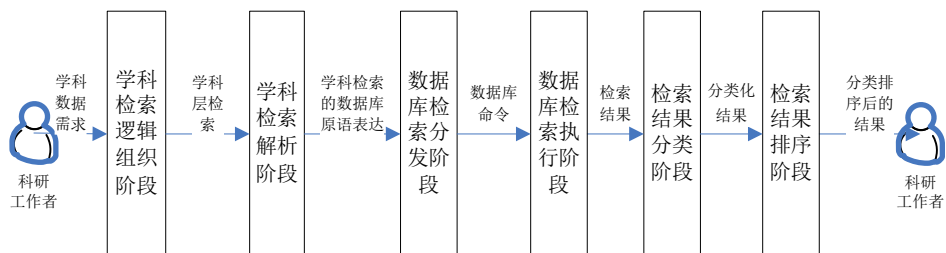


图 6-11 数据资源发现的数据流图

资源发现服务（见图 6-12）主要分为 3 个子服务：基于语义检索的资源发现、基于核心属性的资源发现、基于目录导航的资源发现。

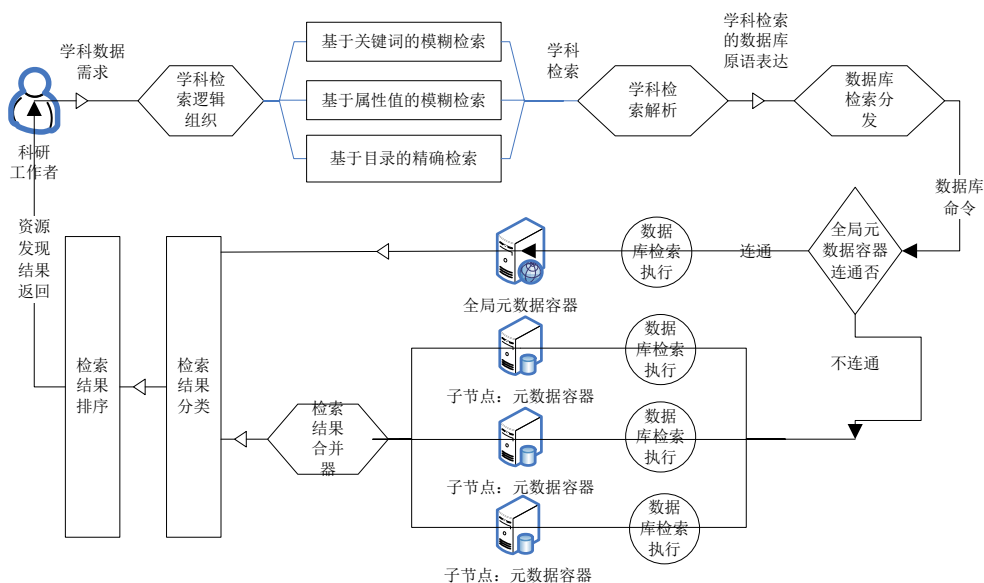


图 6-12 资源发现服务设计

基于语义检索的资源发现，是指以全部核心描述属性的属性值为匹配对象，以“关键词匹配正确”为标准的数据资源检索。基于关键词的模糊检索，是一个人机交互检索的过程。

基于属性值的精确检索：面向属性值的检索指以核心描述属性的物理属性值为驱动的数据检索。主要包含：

- 基于观测物理要素的检索、
- 基于观测平台的检索、
- 基于观测仪器的检索、
- 基于空间观测区域的检索、
- 基于观测时间的检索、
- 基于数据类型的检索。

基于目录的精确检索，指采用物理数据分类分面的思想，针对物理数据的建立完整的分类索引，通过树型多叉查找法分级分层次进行导航式资源发现。该服务由两部分功能组成：①目录树的自动生成和更新；②目录导航的结果范围锁定和检索。目录树的自动生成，指系统依据分类的类型，比如基于学科或者台站等，自动的从全局元数据容器的核心描述属性中，提取出学科目录树的结构，表达到

网页上。目录树的自动更新，指系统自动检测数据资源的变动和更新，并把变动和更新映射到目录树的变化上。目录导航的结果范围锁定和检索，指用户通过目录式的浏览，圈定自己所要的资源后，可以执行两类操作：①直接请求下载该类的数据；②在该圈定的资源范围内，转入基于关键词的精确检索，利用关键词检索进行更细微的范围圈定。

当一个检索请求发出，首先利用“检索解析器”从“学科级别检索原语”解析为“数据库级别检索原语”。此时，可以执行检索的有两个元数据库，一个是全局元数据容器，另一个是所在节点的审核通过元数据容器。如果“数据库级别检索原语”在全局元数据容器执行，检索结果是面向空间科学虚拟观测台的所有分布式资源，是全局的；如果“数据库级别检索原语”在所在节点的审核通过元数据容器执行，检索结果是面向本地节点的，是局部的。系统检索的策略有 2 点。第一，优先投递“数据库级别检索原语”到全局元数据容器，即全局检索结果为优先级最高的结果；第二，如果因为网络状况或者其他问题，导致无法获得“全局元数据容器”的检索结果，则退而求其次，向所在节点的“审核通过元数据容器”投递“数据库级别检索原语”，即本地检索结果为候选结果。

6.3.3 资源存储与访问服务设计

为提高空间科学虚拟观测台的数据资源融合的灵活性，我们采取数据资源描述和数据资源存储访问剥离的思想，摒弃传统数据资源描述和数据资源存储访问必须相捆绑在一地的做法^[44]。因此，资源的加入和发布只需要资源提供者按照公共数据模型进行标准化描述，并通过资源注册服务描述注册到网格系统即可，用户可以任何地方注册数据资源，而注册点不必是数据资源存储的所在地。

资源存储访问服务（见图 6-13），需要面向存储、监控和访问提供不同的功能。

（1）面向存储

存储系统是数据网格的基础设施之一。空间科学 VO 依靠存储系统，实现数据资源的建立、删除、读、写、控制等底层功能，并通过封装和组合底层功能对外提供访问接口。因而，资源存储服务，将以数据节点的方式，为数据网格提供存储力，并解决容错性和健壮性问。

(2) 面向监控

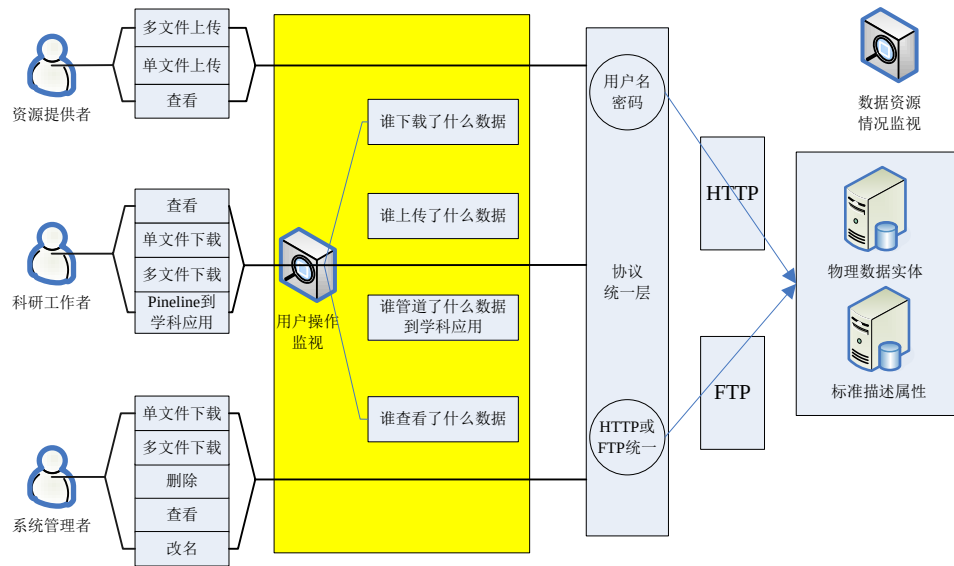


图 6-13 资源存储与访问服务设计

资源存储与访问服务还将为系统监控提供支持（见图 6-14）。利用该服务，系统可以记录谁下载了什么数据、谁上传了什么数据、当前的数据资源状况，并为系统监控服务提供历史记录查询接口。

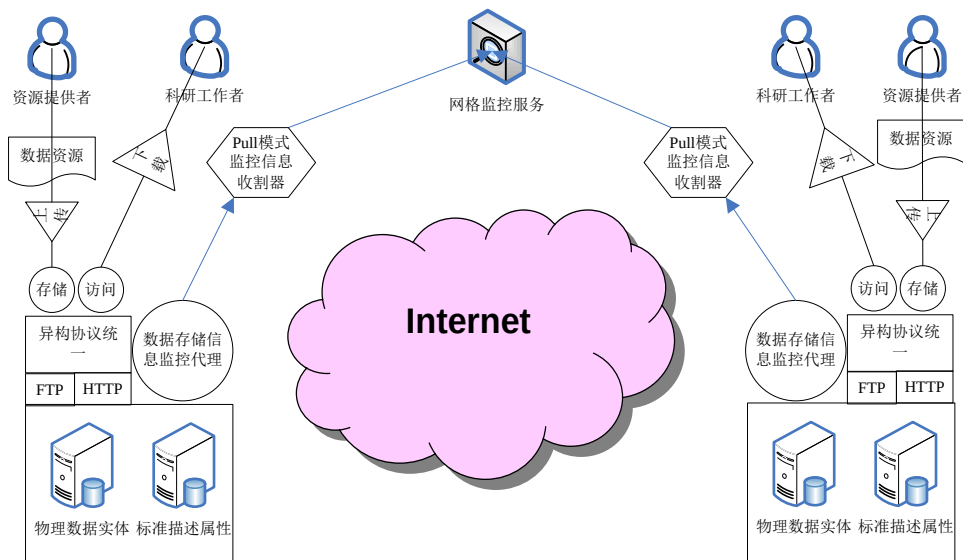


图 6-14 资源访问服务与监控服务的耦合

(3) 面向访问

首先，资源访问服务需要解决访问协议异构的问题。在实际数据因公系统中，不同的数据存储方案将会采用不同数据的访问协议（FTP/HTTP 等）。因而，在访问不同的存储系统时，如果不在体系结构上进行统一，就不得不面临一个尴尬

的事实——必须在多种访问协议中频繁切换。然而，这将会降低在不同存储系统之间传输数据的效率，增加用户使用的复杂性。因而，需要在底层传输协议上，增加一个抽象层，以统一不同的数据访问协议。它实质上是一个新的访问协议，提供了访问数据资源所必须的接口，使得其他服务不需要关注访问协议异构的问题，最终实现透明访问。

其次，利用资源访问服务，解决数据资源访问安全机制的异构性问题。通常，部分资源提供者，往往在访问协议层上增加访问认证机制，如：传统的用户名密码安全认证机制，证书认证机制等。因而，资源存储与访问服务，应该把不同的资源节点独立的安全访问认证机制统一起来，向网格服务屏蔽底层不同的安全访问机制。

第三，利用资源访问服务，解决从物理数据到学科应用的 Pipeline 的问题。我们知道，科学工作流的提出，主要是为了解决数据资源访问服务和学科应用服务之间的关系问题，即如何在完成学科资源检索需求后，按需把物理数据管道到学科应用服务进行计算、分析、可视化等。这里的 pipeline，不是指进程通信层次的 pipeline，而是软件层次的 pipeline。

综上，资源存储与访问服务的意义在于，把异构的访问协议和访问安全机制进行内部统一化，以支持网格系统的透明访问，同时提供单文件和多文件批量下载等类型的应用服务，提高数据访问的便捷性。更重要的是，实现数据提供者和 VO 系统的分离，是保证 VO 低耦合性的基础。

6.3.4 元数据收割同步服务设计

在空间科学虚拟观测台中，由于其分布式体系，容易导致当某一网格节点的数据资源发生变更时，如新数据资源发布、旧数据资源删除等，系统其他节点对于数据资源变更的同步严重滞后。然而，要实现全局资源的快速发现，就要求系统具有自组织能力，能够实时同步全局的数据资源分布。为此，我们提出了元数据收割同步服务。

收割的同步性表现在两方面：及时性和一致性，数据提供者对元数据信息做了增删改后能够及时同步到服务提供者，保证两端元数据一致。综合考虑及时性和一致性的要求，我们设计了一种基于推拉式的元数据收割模式（见图 6-15）。

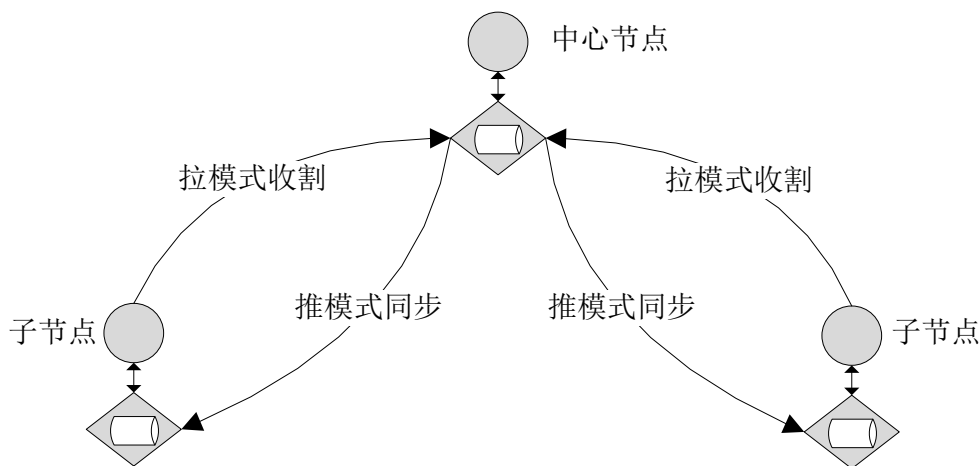


图 6-15 推拉式的数据收割同步策略

拉（pull）模式主要满足两方面的需求：①保障数据更新的实时性。若数据提供者对注册新的数据资源描述或者修改更新已有数据资源描述，系统以固定的时间间隔，自动检测各子节点的元数据的更新和变动，以拉模式收割更新的元数据到全局元数据容器。②更新的数据资源描述审核集中化。通过拉模式，系统把更新的数据资源描述收集到全局元数据容器，由系统管理员进行统一的审核。

推（push）模式主要为了保障子节点的资源描述更新后，其审核结果与子节点的实时同步。更新的资源描述信息审核通过后，该部分数据资源描述信息的状态从“待审核”转移到“已审核”。依据其来源地，系统自动把该部分资源描述信息，通过推模式重新推送回到对应子节点的审核通过元数据容器，以使得子节点实时同步资源的审核结果，保证数据的实时发布。

“拉模式收割”和“推模式收割”的本质区别是：前者同步的是全局的描述信息，后者同步的是全局的描述信息的审核状态变更。

除此以外，我们还需要一个收割协议，用以指明元数据格式和物理含义、收割对象、收割时间跨度、收割频率、收割的触发条件、收割完成后的存放容器。

空间科学 VO 以开源关系型数据库 MySQL 为基础，构建元数据同步收割服务（见图 6-16）。由于服务层上的资源注册、审核等业务流程最终总会转化为对资源层上 MySQL 数据库的具体操作。为此，我们在主节点的服务层和资源层之间，对 SQL 执行语句进行监控、截取和分析，记录导致数据库变动的增加、删除、更新等三类修改性操作。子节点的守护进程则通过定时从主节点收割数据库变动记录并在本地执行的方式，最终实现同步。这种同步方式，是在各节点的数

数据库结构一致的前提下实现的。它要求空间科学 VO 具有统一的数据管理模型，因而具有一定的限制性。利用该思路，我们绕过数据视图的关联操作，规避数据库的实体与实体、实体与属性间的一对多、多对多等复杂的主外键关系，使得数据库变动的监控和记录在实现上具有可行性，在逻辑上具有完备性。

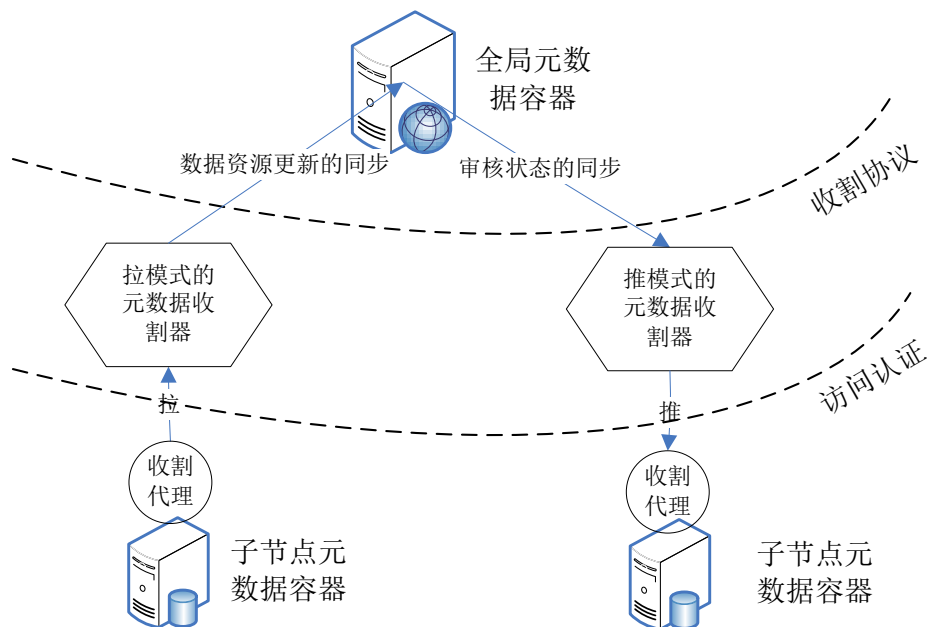


图 6-16 数据收割服务设计

6.3.5 支持一站式访问的单点登录服务设计

一站式单点登录（SSO, Single Sign-on, 见图 6-17）是指网格用户在一次登录系统后，即可任意访问网格中任何节点的数据和服务，而不需要重新登录。

空间科学虚拟观测台拟采用基于票据的单点登录模式，其基本思想是：用户登录网格系统时，输入用户名和密码，系统把用户信息写入到用户浏览器的 cookies 里，在浏览器关闭之前或者限制时间之前，用户信息都存储在 cookies 里，当用户切换访问另外的节点的数据和服务时，系统自动读入用户浏览器 cookies 里的用户信息，完成登录验证，而这一切对于用户是透明的，用户在感觉上是一站式的登录。

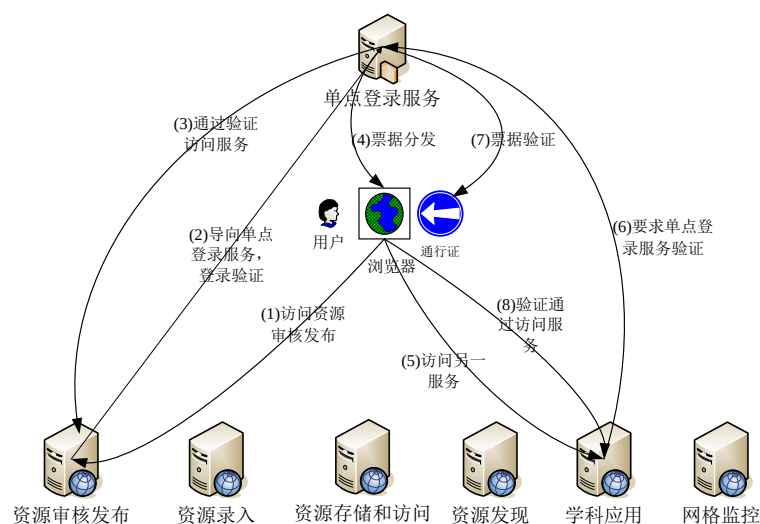


图 6-17 基于票据的单点登录服务设计

如图 6-18 所示，当用户第一次访问网络节点 1 的时候，因为还没有登录，会被引导到认证系统中进行登录；根据用户提供的登录信息，认证系统进行身份校验，如果通过校验，则返回给用户一个认证的凭据，该凭据是保存在用户的 cookie 中；用户再访问其他应用的时候，单点登录服务器会主动获取这个认证凭据，然后进行校验，如果通过校验，用户就可以在不用再次登录的情况下访问应用站点二和应用站点三。基于“ticket”的单点登录机制，使得多个系统共享一个身份认证系统，便于用户在多个应用站点之间自由穿梭。

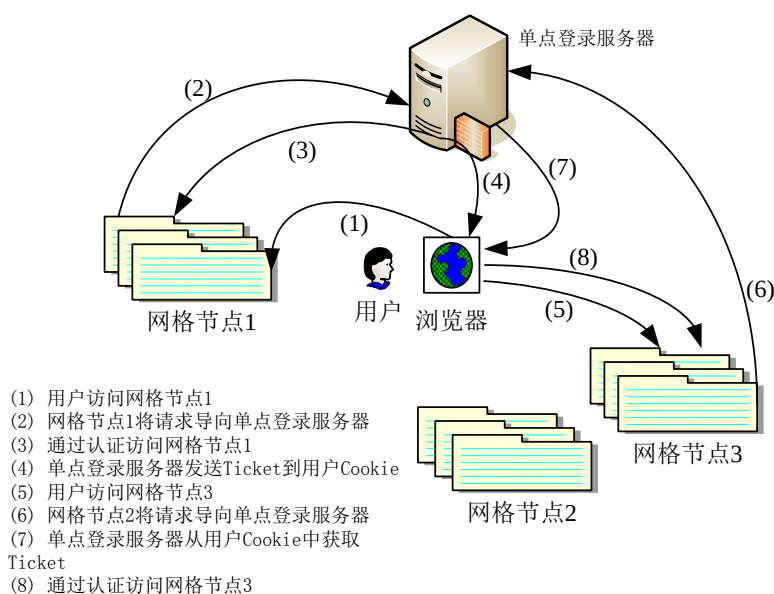


图 6-18 单点登录流程示意图

6.3.6 监控服务设计

网络监控对空间科学虚拟观测台的管理和决策是至关重要的,通过监控可以及时发现数据、存储资源和服务的使用率,分析系统性能,及时发现故障,提高网格系统服务质量,挖掘使用网格节点上用户的分布、数据使用率等信息,从而找到用户所关注的热点数据、热点服务,提高数据服务的资源的利用率,为进一步优化数据、服务和资源提供决策依据。

空间科学虚拟观测台的网格监控服务(见图 6-19)主要针对数据、服务、用户和系统资源等 4 类对象进行监控,并对系统日志和用户行为记录进行统计分析。依据监控对象的不同,网络监控主要分为 4 类监控:①数据监控,指监控是新的数据资源加入、类别数据资源的总量、节点数据资源的总量、数据资源的下载频率、数据资源的检索频率等。②服务监控,指监控服务的使用率、服务的可用性。③用户监控,指监控用户的操作,包括登录登出、个人资料修改、注册数据资源、检索数据资源、下载数据资源、访问学科应用服务等。④硬件监控,指监控 CPU 的使用情况、存储空间的使用情况、网络流量、系统访问负载。

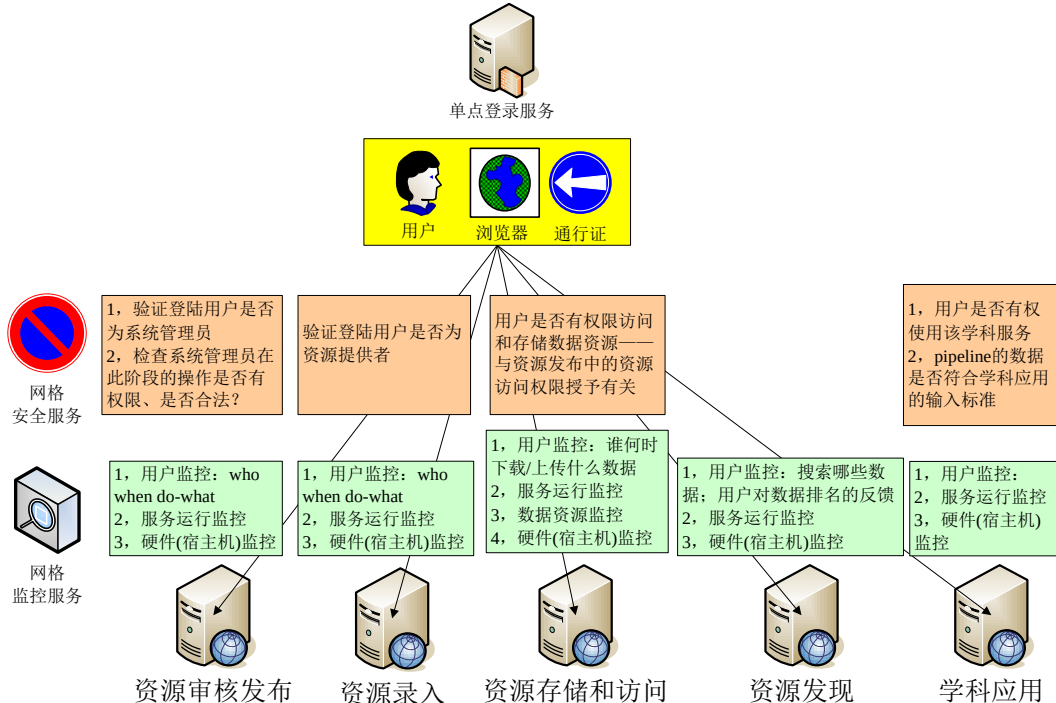


图 6-19 监控服务设计

6.3.8 安全服务设计

网络安全是空间科学虚拟观测台 7*24 小时运行的基本保障。网络安全的核心关注点主要在四个方面：用户访问的身份认证，用户访问的权限授予，数据检索和访问权限的限制，服务访问和使用的权限限制。

用户访问的身份认证，主要包含 4 点要素：①身份认证的统一性，即认证方对用户的认证结果是获得数据网格中所有节点的信任的；②身份认证信息的完备性，即用户提供的身份信息不但充分，而且能够唯一正确地映射到用户数据库的对应网格用户；③身份认证的不可伪造性，即他人无法或者需要付出更大的代价才能枚举或破解出正确的身份认证信息。

用户访问的权限授予，指依据用户角色的不同，网格系统为用户设置其可访问的数据和服务范围，以及访问的深度。比如：能否检索某类资源，返回结果后能否下载该类资源。

数据检索和访问权限的限制，指依据数据资源提供者和数据网格运营单位的需求，对数据资源的检索和访问权限进行设限，以确保特定数据资源的特定共享原则得以满足。

服务访问和使用权限的限制，指依据服务提供者的需求，对服务的开放访问使用所面向的对象进行设置。

6.3.9 学科应用服务设计

科研工作者需要从主题库查询并获取所需要的大量的科学数据，最终使用这些数据进行计算和分析。传统的方式是将这些数据先下载下来再进行计算分析。若数据量较大时，下载这些数据所需的时间较长，此外，科研工作者或许没有足够的计算资源进行分析处理，这些数据又会被上传到计算中心进行处理，这种方式导致了极低的科研效率。通过科学数据网格可以集成数据、存储和计算资源为用户提供高效的科研工作平台

(1) 基于空间事件关联的数据发现服务

传统数据应用系统基于数据属性（载荷、仪器设备等）进行数据发现。但是，这种“实体-属性”式的关联与科研的思维方式相去甚远。依据学科知识，对数据进行因果溯源和关联分析，才是找到新问题的关键。在空间科学领域，空间天

气事件（日冕物质抛射、磁场扰动、质子事件等）与卫星载荷数据的因果时序关联，可以分为3类：①先兆特征型；②伴随发生型；③后续效应型。为此，我们把空间天气事件在时序区间、观测平台、载荷类型上与数据的关联关系视为知识库，并在其监督下进行数据发现。

空间科学 VO 的资源发现服务（见图 6-20）属于半监督检索系统，由查询构造引擎、查询结果引擎和监督知识库构成。“查询构造引擎”在“监督知识库”的指导下，把学科数据需求转化为符合 SQL 标准的可执行命令；“查询结果引擎”对原始的 SQL 结果对象进行解析、分类、排序后，结构化地展现结果。图 4 展现了数据发现的流程：①在“空间天气事件关联知识库”监督下，把用户的学科数据需求描述为学科查询。②依据“学科数据模型”，构建一个逻辑上的数据视图。该视图为学科查询涉及的所有数据，并把“学科查询”转化为“数据视图查询”。③根据数据库表属性的关联关系，把视图查询分解为表实体的关联查询，并构造符合 SQL 语法规则的语句；④提交 SQL 语句到数据库引擎；⑤获得 SQL 结果对象；⑥对结果分类；⑦对结果排序并展现给用户。

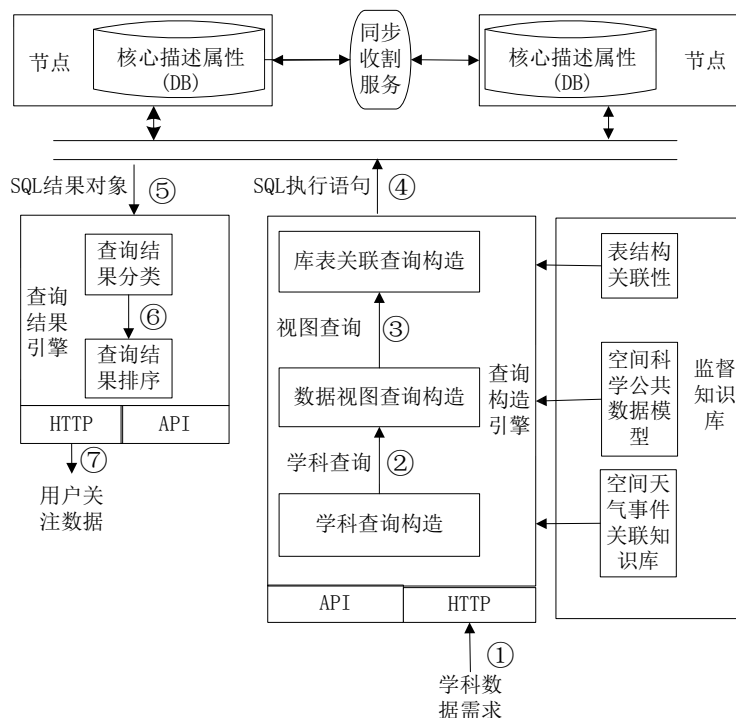


图 6-20 基于空间天气事件的资源关联发现

空间科学数据资源的一个重要特点是观测数据的相关性强。针对一个研究对象，通常从不同的观测平台、不同的观测手段，不同的观测角度，不同的物理属

性等去进行观测和记录。从一个研究对象引出的空间科学事件，影响着不同空间观测区域和观测对象的物理特性，并导致相应的变化。然而，当前的科学数据应用中，却鲜有利用空间事件的相关性，对全局的观测数据进行数据发现。基于空间科学事件关联的数据发现正是基于这个出发点，以空间科学事件对观测数据的影响为导向，紧密依据空间科学事件与数据的时间联系和物理联系，实现数据资源的关联发现，而非普通检索。

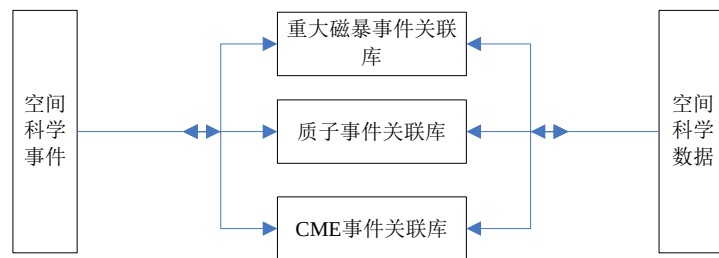


图 6-21 空间科学事件关联发现三要素

如图 6-21 所示，为建立某一空间科学事件与空间科学数据的关联，数据网格系统将会维护三个部分的数据表达，分别为：空间科学数据、空间科学事件编目和事件关联库。三部分数据表达中的核心是事件关联库，通过事件关联库，实现事件和数据的关联映射关系。

在基于空间天气事件的资源关联发现方面，空间科学 VO 以太阳质子事件为切入点，着力维护三类数据表达：空间科学数据索引库、太阳质子事件编目库和太阳质子事件关联库。在三类数据表达中，最核心的是太阳质子事件关联库，它作为“桥梁”，构建起从质子事件到科学数据的关联映射关系。该关联关系主要分为四部分：①质子事件；②与质子事件对应的观测平台（科学卫星或地面台站）；③与质子事件对应的观测设备；④质子事件与科学数据在时间属性上的关联类型（先兆特征型、伴随发生型、后续效应型）；⑤质子事件与科学数据在时间属性上的映射跨度。针对某一质子事件，系统首先依据其所关联的观测平台和观测设备，在空间科学数据索引库中确定具体的数据范围。如果是先兆特征型关联，则以质子事件爆发的时间点为终点进行前溯；如果是伴随发生型关联，则以质子事件最大通量的时间点为中心，进行中心时间区间截取；如果是后续效应型关联，则以质子事件结束的时间点为起点进行后推，从而最终确定关联数据。

(2) 美国海军实验室 MISIE 模式计算及其科学可视化服务

美国海军实验室 MSISE-00 Model 2001 全球大气密度和温度模式，是一个经验模型，主要用于分析和模拟从 10 km 到 45 km 空间的大气成分（包括：O₂、N₂、He、Ar 等）密度或温度（Neutral Temperature、Exospheric Temperature）。该服务由大气模式计算和大气密度可视化两个核心模块组成（见图 6-22）。

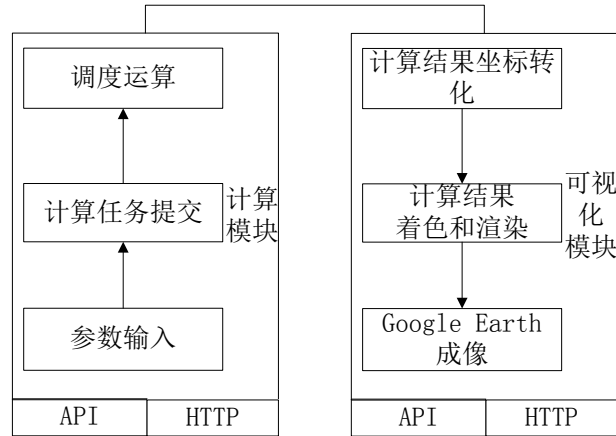


图 6-22 MISIE 大气模式计算及科学可视化服务

在模式计算阶段，用户可以通过模式计算交互界面设置和输入大气成分、年份、月份、日期、时间类型（世界时或地方时）、计算坐标系（地理坐标或地磁坐标）、高度等参数，MSISE 模式自动从各节点查询并载入计算参数和对应的物理观测数据，计算并反馈计算结果。用户可以在线浏览下载计算结果。

在可视化阶段，系统自动载入大气模式的计算结果文件，获得某大气成分在全球不同经度、维度和高度上的密度，并且在 RGB 彩色坐标中，把大气密度从冷色调到暖色调，映射为一个具体的颜色，再用插值法对没有计算大气密度的区域进行平滑，最终作出全球分布图，从而进一步揭示不同大气成分的全球密度分布规律。

(3) NASA SPASE 元数据标签在线解释服务

空间物理档案搜索和抽取（SPASE）标准，是美国宇航局提出的面向日地物理学的元数据标准。SPASE 的提出主要有 4 个目的：①基于公用元数据语言（XML），实现空间科学数据环境的数据检索和抽取的加速；②定义和维护空间太阳物理学的元数据标准，以支持空间和太阳物理数据互操作；③使用 SPASE 元数据标准为卫星观测数据创建标准数据描述；④为数据工程师和物理研究者提供 SPASE 元数据标准的辅助工具。

SPASE 元数据标签解释服务，通过读取卫星数据的元数据标签文件（XML），

按照 SPASE 标准, 对 XML 内容遍历, 在遍历过程中同时进行语法检查和语义解释, 最终把卫星数据的描述信息用一个通俗易懂的方式展示给科研工作者, 以方便科研人员读懂卫星数据的来源、仪器、设备等相关信息。

(4) CDF/FITS/netCDF/DAT/HDF/CDFML 在线转换服务

空间科学数据的表达和存储方式多种多样, 目前主流常用的数据格式有 8 种之多, 分别是: CDF、netCDF、FITS、ASCII、CDF Skeleon Table、CDFML、HDF4、HDF5 等。为加快和方便空间科学数据在不同格式标准之间进行快速的交互和转换, 空间科学虚拟观测台拟以 Web Service 的方式, 在线提供“空间科学数据云转换工具”, 使得空间科学虚拟观测台用户无论何时何地都能轻松获取丰富的数据格式转换服务, 而不需要本地安装和运行。

在线转换服务目前主要提供: CDF-netCDF、FITS-CDF、netCDF-CDF、HDF4-CDF 四种数据格式转换服务, 同时支持对科学数据文件的本地或跨节点远程的转换, 并且与虚拟观测台中的数据发现服务链接起来, 实现从数据发现服务到数据转换服务的自动传输和管道。

(5) CDF 在线阅读服务

CDF (Common Data Format) 是一种自描述的数据格式, 被空间科学界广泛用于多维或大规模数据的存储和互操作应用。CDF 属于 ASCII 文件, 因而通常需要研究人员用专用软件打开, 才能阅读数据内容。

CDF 在线阅读服务实现 CDF 文件阅读的在线化。科研人员可以通过该服务直接在浏览器中阅读 CDF 数据内容, 而不需要安装单机版软件。该服务也支持对科学数据文件的本地或跨节点远程的阅读, 并且与数据发现服务链接起来, 用户找到合适的文件后, 点击阅读, 即可阅读数据内容, 系统后台自动实现数据在两个服务间的管道与传输。

6.4 系统部署

基于上文所阐述的体系结构, 我们实现了空间科学 VO 的原型。如图 6- 23 所示, 空间科学 VO 由 1 个主节点和 4 个学科子节点组成。其中, 主节点、太阳、地磁和中高层大气节点分别位于中国科学院空间科学与应用研究中心、中国科学院国家天文台、中国科学院地质与地球物理所、以及中国科学技术大学, 且都统

一部署了资源注册、发现、存储与访问、同步收割、透明访问、用户管理、安全监控和数据格式化管道等服务。图 6- 24 展示了空间科学虚拟观测台的树式主从管理拓扑。

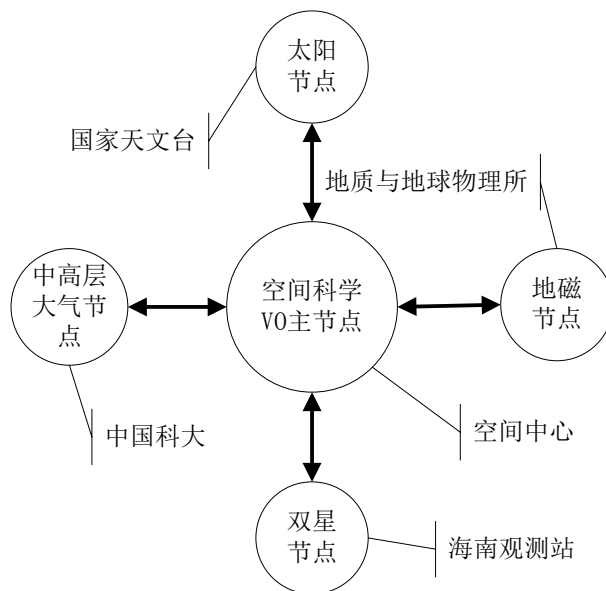


图 6- 23 空间科学虚拟观测台节点分布

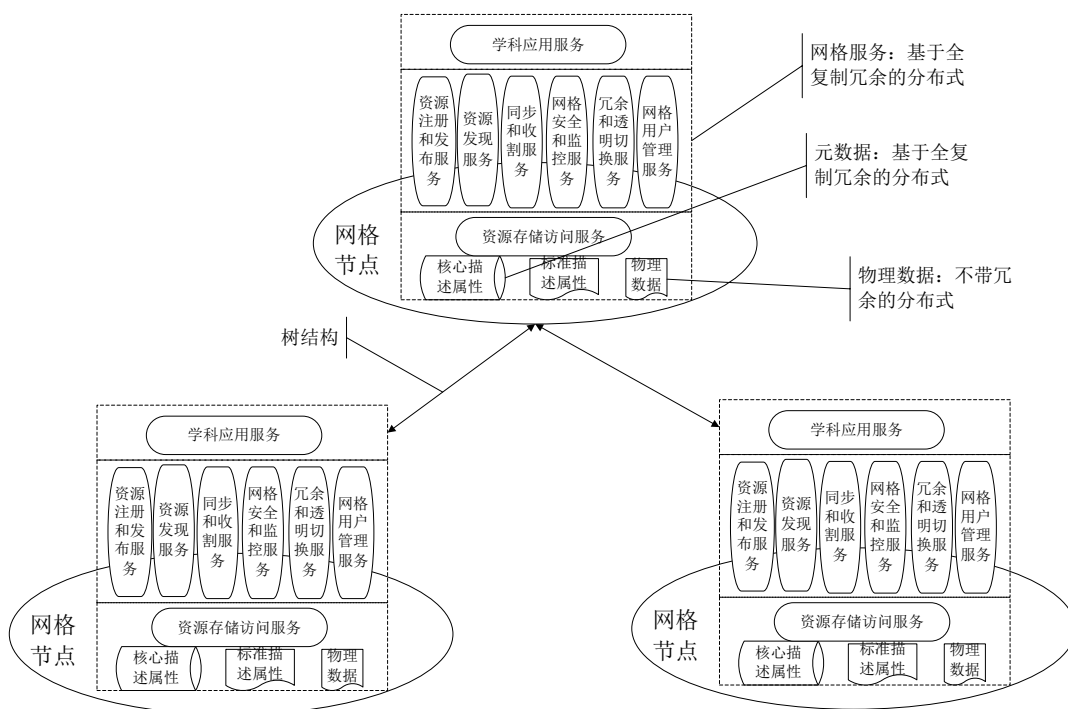


图 6- 24 空间科学虚拟观测台拓扑结构

6.5 科学应用

为解决数据资源与学科应用资源的互操作和应用集成问题,空间科学虚拟观测台特别针对数据应用集成和研究热点的需求,提出并设计了空间科学数据服务、空间天气模式计算服务、应用工具在线服务、空间天气事件关联数据发现服务共四类学科应用资源服务中间件,并与基础支持服务各类中间件实现了互操作。

6.5.1 空间科学数据服务

空间科学数据服务中间件提供数据目录查询、数据属性查询、模糊查询等多种简单、易用的数据检索方法(见图 6-25),科研用户可在任意节点输入查询条件,空间科学虚拟观测台会自动将各节点符合条件的数据集反馈给用户,每个数据集提供了详细的数据描述信息和辅助数据文档,并提供数据在线阅读、数据下载、数据可视化等服务,科研用户可随机选择服务,了解数据、阅读数据、浏览数据和下载数据。

目前,空间科学虚拟观测台以在线和离线方式为国家自然科学基金、中国科学院创新性项目在空间物理、空间环境研究与应用研究等多个前沿性课题提供数据检索、定位、打包、压缩、传输、获取等常规数据服务;利用自身在学科数据规范、数据库技术、数据档案管理等方面的优势,为国家重大项目提供技术集成服务;为国家自然科学基金、中国科学院创新性项目在空间物理、空间环境研究与应用研究等多个前沿性课题提供了离线数据服务。

The image shows a web-based data search interface. On the left, there is a sidebar with a tree view of categories: '日地空间物理' (Solar and Geospace Physics), '行星科学' (Planetary Science), and '国际镜像库' (International Mirror Library). The main content area is titled '数据检索' (Data Search). It contains several search criteria: '区域' (Region) with a dropdown menu, '观测设备' (Observation Equipment) with a dropdown menu, '观测平台' (Observation Platform) with a dropdown menu, '卫星' (Satellite) with a dropdown menu, '物理要素' (Physical Elements) with a dropdown menu, '关键字' (Keywords) with a text input field, and '时间范围' (Time Range) with two date pickers. A 'Search' button is located at the bottom right of the search criteria section. Below the search criteria, there is a section titled '检索条件容器' (Search Condition Container) with '清除' (Clear) and '检索' (Search) buttons.

图 6-25 基于核心属性、语义与分类导航的数据发现

检索结果

- ACE Cosmic Ray Isotope Spectrometer (CRIS) Level 2 Data in CDF F...
- ACE Electron Proton Alpha Monitor (EPAM) Level 2 Data in CDF Form...
- ACE Magnetometer(MAG) Level 2 Data in CDF Format
- ACE Solar Isotope Spectrometer (SIS) 1-Hour Level 2 Data in CDF F...
- ACE Solar Wind Electron Proton Alpha Monitor (SWEPAM) Level 2 Dat...
- BBSO Full-disc Halpna Original Data in JPG Format
- Beijing Ionospheric Characteristic Parameter Data in TXT Format
- Beijing Ming Tomb Geomagnetic Field Data in TXT Format
- CAO Full-disc Halpna Original Data in JPG Format
- CE-1 Microwave Sounder Brightness Temperature Data in DAT Format

Page 1 of 17 Displaying 1 - 10 of 170

ACE Cosmic Ray Isotope Spectrometer (CRIS) Level 2 Data in CDF Format

VSSO Node: Space Science Master Node
Dataset Producer: ACE
Observe Equipment: Particle Detector
Observe Area: [Helosphere.Near Earth]
Observe Element: [Interplanetary High energy particles]
Metadata Document

	FILE NAME	START TIME	END TIME	FUNCTIONS
☐	cris_data_1hr_2240.cdf	1997-08-14 00:00:00	1997-09-09 00:00:00	Get Read Translate
☐	cris_data_1hr_2241.cdf	1997-09-10 00:00:00	1997-10-06 00:00:00	Get Read Translate
☐	cris_data_1hr_2242.cdf	1997-10-07 00:00:00	1997-11-02 00:00:00	Get Read Translate
☐	cris_data_1hr_2243.cdf	1997-11-03 00:00:00	1997-11-29 00:00:00	Get Read Translate
☐	cris_data_1hr_2244.cdf	1997-11-30 00:00:00	1997-12-26 00:00:00	Get Read Translate
☐	cris_data_1hr_2245.cdf	1997-12-27 00:00:00	1998-01-22 00:00:00	Get Read Translate
☐	cris_data_1hr_2246.cdf	1998-01-23 00:00:00	1998-02-18 00:00:00	Get Read Translate
☐	cris_data_1hr_2247.cdf	1998-02-19 00:00:00	1998-03-17 00:00:00	Get Read Translate
☐	cris_data_1hr_2248.cdf	1998-03-18 00:00:00	1998-04-13 00:00:00	Get Read Translate
☐	cris_data_1hr_2249.cdf	1998-04-14 00:00:00	1998-05-10 00:00:00	Get Read Translate

图 6- 26 数据检索与定位

6.5.2 基于空间事件关联的数据发现

Search Your Event

Space Weather Event List: Time Scale: Start: End:

Space Weather Events

Start Time	Max Time	Proton Flux	Associated CME	Importance	Region Location	Region Number	Alias
2006-12-13 03:10...	2006-12-13 09:25...	698	Halo /13 0254	X3/4B	S05W23	10930	
2006-12-06 15:55...	2006-12-07 19:30...	1980	Halo	X9/2N	S07E79	10930	
2005-09-08 02:15...	2005-09-11 04:25...	1880	East /07 1723	X17/3B	S06E89	10808	
2005-08-22 20:40...	2005-08-23 10:45...	330	Halo /22 1730	M5/1N	S12W60	10798	
2005-07-27 23:00...	2005-07-29 17:25...	41	Halo /27 0454	M3	N11E90	10792	
2005-07-14 02:45...	2005-07-15 03:45...	134	Halo /13 1430	M5	N10W80	10786	
2005-06-16 22:00...	2005-07-17 05:00...	44	West /16 2003	M4	N09W87	10775	
2005-05-14 05:00...	2005-05-15 02:40...	3140	Halo /13 1722	M8/2B	N12E11	10759	
2005-01-16 02:10...	2005-01-17 17:50...	5040	Halo /15 2306	X2	N15W05	10720	
2004-11-07 19:10...	2004-11-08 01:15...	495	Halo /07 1706	X2	N09W17	10696	

Page 1 of 23 Displaying 1 - 10 of 22

图 6- 27 空间天气质子事件检索

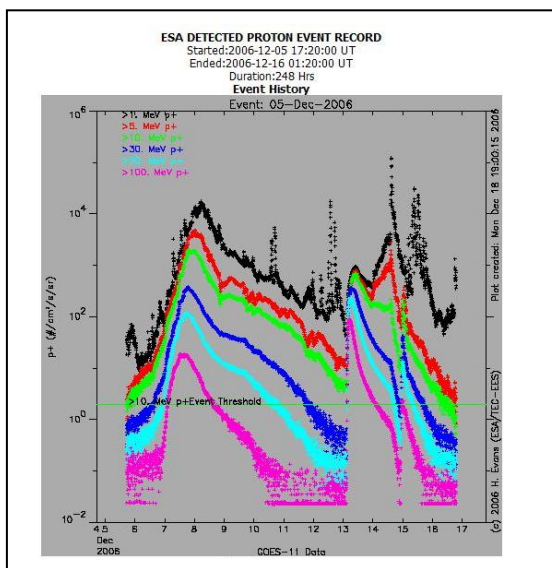


图 6- 28 质子事件记录可视化分析结果

空间天气事件呈现的是一个从太阳到行星际到地球空间的动态因果链关系，一个监测到的空间天气事件，通常会牵引出一系列监测对象物理特征的突变，并直接反映为监测数据的变化。

空间科学虚拟观测台提供了太阳质子事件和 CME 事件两类数据发现服务中间件，服务基于事件与数据的因果时序关联关系进行识别，即先兆特征、伴随发生和后续效应三类。科研用户可以通过事件编目，查询到事件的特征信息（见图 6-27），可视化过程分析（见图 6-28），并从各节点通过获取到与事件相关的数据资源信息（见图 6-29）。

目前，基于事件关联的数据发现中间件，为空间环境预报、空间环境月刊等特殊用户团体提供空间天气预报的专项数据服务。

Event Relative Data				
Relative Dataset	Relative File	Start Time	End Time	Download
ACE_CRIS_1H_127	cris_data_1hr_2366.cdf	2006-12-07 00:00:00	2007-01-02 00:00:00	Get
ACE_EPAM_1H_127	epam_data_1hr_2366.cdf	2006-12-07 00:00:00	2007-01-02 00:00:00	Get
ACE_MAG_1H_127	mag_data_1hr_2366.cdf	2006-12-07 00:00:00	2007-01-02 00:00:00	Get
ACE_SIS_1H_127	sis_data_1hr_2366.cdf	2006-12-07 00:00:00	2007-01-02 00:00:00	Get
ACE_SWEPAM_1H_121	swepam_data_1hr_2366.cdf	2006-12-07 00:00:00	2007-01-02 00:00:00	Get
DRAO_F107_1D_60	2006.ADJ	2006-01-01 00:00:00	2006-12-31 00:00:00	Get
KYOTO_DST_1H_50	dst2006	2006-01-01 00:00:00	2006-12-31 00:00:00	Get
NGDC_AP_3H_75	ap2006.bt	2006-01-01 00:00:00	2006-12-31 00:00:00	Get
NGDC_KP_3H_75	kp2006.bt	2006-01-01 00:00:00	2006-12-31 00:00:00	Get
SIDC_SSN_1D_189	2006	2006-01-01 00:00:00	2006-12-31 00:00:00	Get

图 6-29 质子事件关联数据检索

6.5.3 科学数据的在线可视化

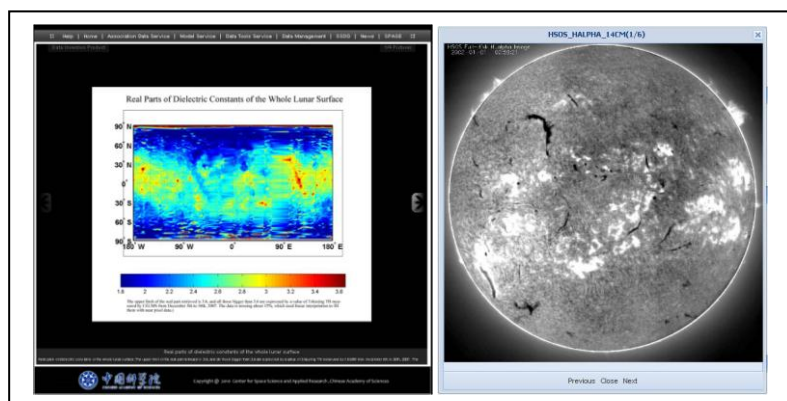


图 6-30 嫦娥月球微波数据和怀柔太阳观测数据的在线可视化

空间科学虚拟观测台向科研用户提供基于 IDL、Matlab 等工具的常规二维科学数据在线可视化（见图 6-30），主要分为两类：①对物理要素探测数据的可视化，表现形式为在时间轴上延伸的散点图、柱状图、折线图；②对望远镜拍摄

数据的可视化，主要为太阳跟踪拍摄图像。科研用户对检索结果中的数据点击 View、Plot 等功能后，系统自动调用后台程序，把图片数据加载到前台，并允许用户放大缩小或连续播放。

6.5.4 空间科学数据在线阅读与转换

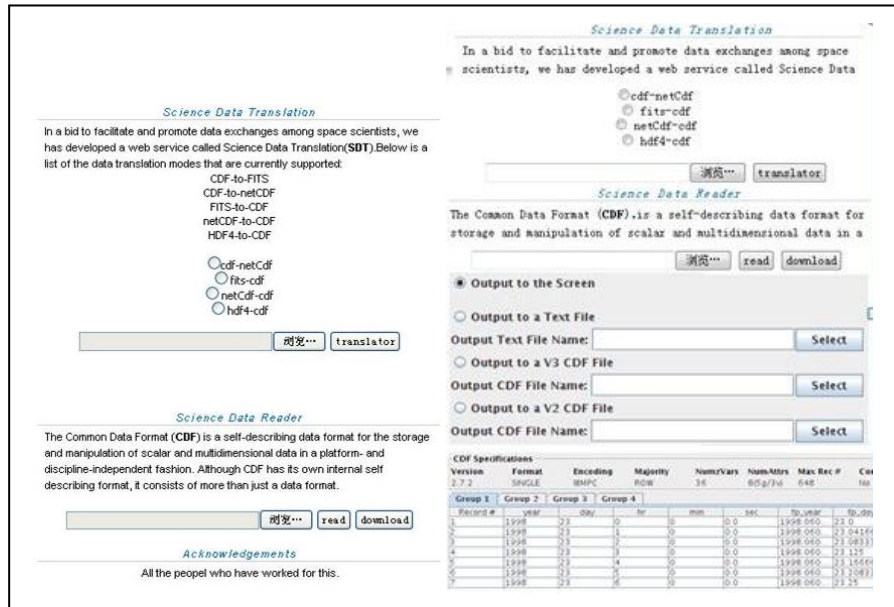


图 6- 31 CDF 在线阅读和多格式互转

针对空间科学数据存储格式多样的特点，空间科学虚拟观测台提供有 CDF-to-FITS、CDF-to-netCDF、FITS-to-CDF、netCDF-to-CDF、HDF4-to-CDF 的数据文件在线格式转化服务中间件和 CDF 数据文件的在线阅读服务中间件，科研用户可直接将在虚拟观测台上获取的数据进行在线阅读和格式转换，或提交本地数据文件进行在线阅读和转换（见图 6- 31）。

数据在线阅读服务，不仅提供物理观测数据阅读，还提供 SPASE 标签元数据的阅读服务（见图 6- 32），以方便研究人员了解该数据集的卫星平台、仪器设备、观测区域、探测要去、单位标量等辅助信息。

这三种服务，不但支持本地提交数据，也支持以 URL 定位的远程数据阅读和转换，因而，用户既可以处理单机数据，也可以处理 VO 内跨节点的数据。

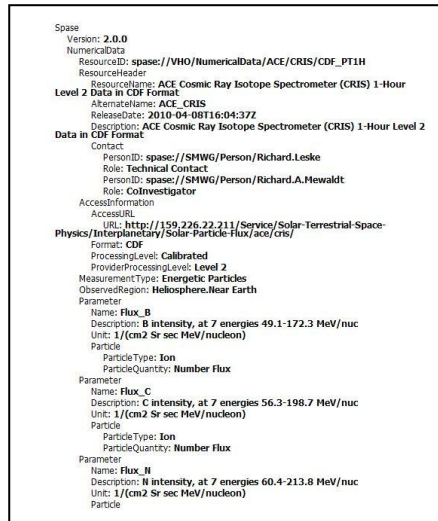


图 6-32 SPASE 元数据标签阅读器

6.5.5 空间科学模式计算及其可视化

为了示范性解决数据与模式计算的互操作问题，空间科学虚拟观测台目前提供了大气 NRLMSISE 模型（见图 6-33）和 IGRF 地磁模型（图 6-34）两类经典空间天气模式在线计算服务中间件。科研用户可通过模型计算交互界面选择计算内容，模式将自动从各节点查询并载入计算参数、计算，并反馈计算结果，用户可在线浏览、下载计算结果。系统同时提供基于 Google Earth 的可视化服务，用户只需安装 Google Earth 可视化插件，模式计算结果即以可视化方式直观展现。

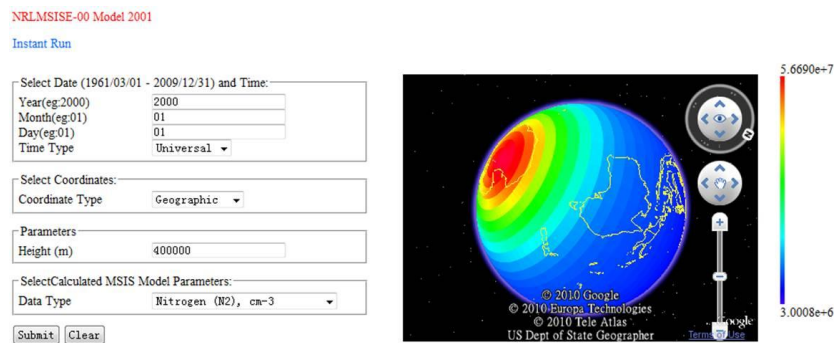


图 6-33 美国海军实验室 MISI 大气模型在线计算及三维可视化

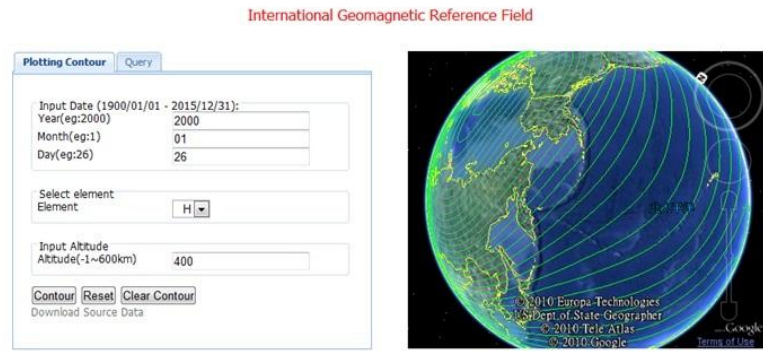


图 6- 34 IGRF 地磁模型在线计算及其三维可视化

第七章 结语

作者在其硕导师的指导下,分别从空间科学 VO 体系结构的理论研究、运行引擎、原型实现和科学应用四方面开展其硕士论题。本硕士论题的主要贡献,体现在如下 5 方面:

1, 面向空间科学 VO 体系结构,在借鉴国内外研究成果的基础上,提出了一种 VO 体系结构的分类方法,并对空间科学领域的典型 VO 体系结构进行分析。

2, 紧密结合我国空间科学应用研究的特点,开展空间科学 VO 体系结构设计理论的探索,详细分析了体系结构设计中的核心问题,提出了空间科学 VO 的基本设计原则和空间科学数据管理模型,然后从概念框架、层次结构、交互协议和服务模型四个方面对空间科学 VO 体系结构进行研究,并进行理论归纳和总结。

3, 针对传统数据应用环境建设中遇到的问题,总结了空间科学 VO 运行引擎的设计动机,然后提出了基于 MVC 设计模式的运行引擎设计方案,并结合硬件发展态势予以改进,最后对其工作流的完备性予以推导和证明。论文把运行引擎划分为框架、控制、视图和模型四个部分,分别进行设计和实现,最终构成具有重用性的空间科学 VO 的运行引擎。

4, 基于空间科学 VO 运行引擎,结合空间科学数据管理模型,设计和实现了空间科学 VO 的分布式数据库,以及门户服务、资源注册服务、资源发现服务、资源存储与访问服务、资源收割同步服务、学科应用服务等核心服务。

5, 结合空间科学 VO 原型,我们分别在空间科学数据服务(目录导航、检索、定位、传输、下载、批量处理等)、基于空间事件关联的数据发现、科学数据在线可视化、科学数据在线阅读和多格式转换、SPASE 标签语义解释、空间科学模式计算及其基于 Google Earth 三维可视化为科研工作者提供具有实用价值的学科应用。

本文面向空间科学的数据特点,对空间科学 VO 做针对性的体系结构设计,并结合“空间科学主题数据库及应用环境建设”项目的实践,解决了 4 个问题:①在三层式的数据模型中,通过对 SPASE 和 PDS 进行核心属性抽取,进一步推动两个学科数据管理模型的统一。②在数据库结构一致的前提下,利用同步收割服务,实现资源信息的延迟性全局同步。③增加数据格式化管道服务,以解决数

据 IO 异构问题, 构建一个包括“数据查询、定位、获取、阅读、转换和可视化”的科学 workflow 雏形, 为模式计算提供支持。④利用“质子事件关联”构建面向科学问题的数据发现。

本文通过“数据格式化管道”解决数据 IO 异构的思路, 对于天文 VO 将有一定的适应性。此外, 通过建立抽取系统, 对“标准描述属性层”的元数据标准封装文件 (VOTable、PDS 标签、SPASE 标签等) 进行属性抽取, 以降低数据清洗成本的思路是值得肯定的。

然而, 本文仍有一些问题值得进一步探讨: ①本文仅在体系结构设计上借助业务流实现资源信息的延迟性同步, 并没有从数据库技术角度根本上解决实时同步的问题。②在三层式的数据模型中, 大部分“核心描述属性 (DB)”是通过“标准描述属性 (XML)”进行属性抽取获得的。但是, 目前系统仅支持静态抽取, 我们将会进一步致力于发展 XML 动态抽取技术, 以支持抽取属性的动态选择以及属性值变化时同步的动态触发。③如何在质子事件关联的基础上, 通过构建更复杂的多空间天气事件的联合关联检索, 如: CME、磁暴、太阳风暴等, 从而提高数据发现的科学价值, 将会是空间 VO 在新科研形势下的一个重点发展方向。

参考文献

- [1] 孙广楷. 谈谈空间科学技术 [M]. 上海科学技术出版社, 1979.
- [2] 汪定雄 关. 探奇揽胜九重天 : 地球物理与空间科学技术 [M]. 华中理工大学出版社, 2000.
- [3] 黄本诚, 马有礼. 航天器空间环境试验技术 [M]. 国防工业出版社. 2002.
- [4] 邸乃庸. 载人航天器 [J]. 太空探索 2006, 8): 36.
- [5] 崔辰洲. 中国虚拟天文台系统设计 [D]. 北京; 中国科学院研究生院, 2003.
- [6] TODD KING J M, RAYMAON WALKER, STEVEN JOY, TOM NAROCK. The architecture of a multi-tiered virtual observatory [J]. Earth Science Informatics, 2008, 1): 8.
- [7] 邹自明, 佟继周. 空间环境预报数据库系统 [M]//中科院空间环境研究预报中心. 空间环境及其应用专题研讨会论文摘要集. 中国广西北海. 2004.
- [8] COMMITTEE N I S. Toward a National Virtual Observatory: Science Goals, Technical Challenges, and Impletation Plan [M/OL]. 2000[http://www.us-vo.org/pubs/files/white_paper.pdf].
- [9] SRIKUMAR VENUGOPAL R B, KOTAGIRI RAMAMOHANARAO. A Taxonomy of Data Grids for Distributed Data Sharing, Management, and Processing [J]. ACM Computing Surveys, 2006, 38(3): 53.
- [10] FLORIAN SCHINTKE A R. Modeling Replica Availability in Large Data Grids [J]. Journal of Grid Computing, 2003, 9.
- [11] ENGBRETSON M. A Space and Solar Physics Data Model from the SPASE Consortium [M]. Version:220. 2011.
- [12] JET PROPULSION LABORATORY C I O T. Planetary Data System Standards Reference [M]. 2009.
- [13] 张彦霞, 赵永恒. 虚拟天文台: 科学、工具及应用 [J]. 天文学进展, 2006, 23(3): 11.
- [14] MICHAEL C. DOCONTA L J O, KEVIN T. SMITH. 语义网: XML、Web 服务和知识管理的未来 [M]. 中国科学技术出版社. 2009.

- [15]何戈, 徐志伟. 数据网格技术研究 [J]. 微电子学与计算机, 2003, 4): 8.
- [16]张兵利, 裴亚辉. 网格体系结构综述 [J]. 中国科技信息, 2008, 21): 2.
- [17]洪学海, 许卓群, 丁文魁. 网格计算技术及应用综述 [J]. 计算机科学, 2003, 30(8): 6.
- [18]周园春, 吴开超, 佟强, 闫保平. 科学数据网格中数据访问中间件的研究 [J]. 微电子学与计算机, 2006, 23(11): 3.
- [19]董晓霞, 吕延杰. 云计算研究综述及未来 [J]. 北京邮电大学学报, 2010, 12(5): 6.
- [20]肖迎元. 分布式实时数据库技术 [M]. 科学出版社, 2009.
- [21]蔡月茹 柳. Web Service 基础教程 [M]. 清华大学出版社, 2005.
- [22]周传德, 秦树人, 尹爱军. 科学可视化理论及智能虚拟显示系统 [M]. 科学出版社, 2007.
- [23]张荣华. 交互式计算机图形学: 基于 OpenGL 的自顶向下方法 [M]. 电子工业出版社, 2009.
- [24]韩培友. IDL 可视化分析与应用 [M]. 西北工业大学出版社, 2006.
- [25]周开利, 邓春晖. MATLAB 基础及其应用教程 [M]. 北京大学出版社, 2007.
- [26]刘健, 林刚华. 构建虚拟太阳天文的中国数据节点 [J]. 天文研究与技术, 2008, 5(3): 7.
- [27]S. G. DJORGOVSKI R W. Virtual Observatory: From Concept to Implementation [J]. ASPCS, 2005, 345(
- [28]HILL F. The Virtual Solar Observatory - A White Paper [J/OL] 2000, <ftp://lasco6.nascom.nasa.gov/pub/vso/wpvso2.pdf>.
- [29]J.B. GURMAN G D, J. HOURCL The Virtual Solar Observatory: Still a Small Box [M]. 2006.
- [30]HILL F. THE VIRTUAL SOLAR OBSERVATORY CONCEPT [M]. 2002.
- [31]Virtual Space Physics Observatory [M]. 2011.
- [32]Virtual Heliospheric Observatory [M]. 2011.
- [33]JOHN F. COOPER R E M, NAND LAL, ADAM SZABO, MATTHEW E. HILL, THOMAS P. ARMSTRONG, ROBERT B. MCKIBBEN, CECIL TRANQUILLE. Virtual Cosmic Ray Observatory (ViCRO); proceedings of the 30TH

INTERNATIONAL COSMIC RAY CONFERENCE, Mexico City, Mexico, F, 2008 [C].

[34]MERKA J, 2008.

[35]Virtual Ionosphere, Thermosphere, Mesosphere Observatory [M]. 2011.

[36]RICHARD PAPPALANGLEY J I P, RALPH D. BUEHRLE. Vibro-Acoustics Modal Testing at NASA Langley Research Center; proceedings of the Proceedings of the CEAS/AIAA/ICASE/NASA Langley International Forum F,1999 [C].

[37]Virtual Energetic Particle Observatory [M]. 2011.

[38]ARVISET C, 2011.

[39]赵仁铃. 基于非技术途径的元数据互操作研究 [J]. 图书情报研究, 2010, 3(3): 4.

[40]中晓娟, 高红. 从元数据映射出发谈元数据互操作问题 [J]. 国家图书馆学刊, 2006, 4):

[41]韩夏, 李秉严. 元数据的互操作研究 [J]. 情报科学, 2004, 22(7): 4.

[42]张宇, 王映辉, 张翔南. 基于 Spring 的 MVC 框架设计与实现 [J]. 计算机工程, 2010, 36(4): 4.

[43]施卫峰. 基于 MVC 的分布式作战应用框架研究 [J]. 指挥控制与仿真, 2010, 32(3): 5.

[44]李令坤, 罗泽, 阎保平. 虚拟天文台中星表数据访问网格服务的设计与实现 [J]. 计算机应用研究, 2006, 2): 4.

硕士期间发表文章目录

期刊论文：

- [1] 傅衍杰, 邹自明, 佟继周, “空间科学虚拟观测台体系结构研究”, 天文研究
与技术, 2011, 第 3 期 (已接收, 待发表)
- [2] 佟继周, 邹自明, 傅衍杰, 郑程, 空间科学 e-Science 应用—空间科学虚拟观
测台 (VSSO), 科研信息化技术与应用, 2011 年第 2 卷 第 1 期: 61-68

会议论文：

- [3] 傅衍杰, 邹自明, 佟继周, 高文健, 面向自主网格的空间科学虚拟观测台体
系结构研究, 第十届科学数据库与信息技术会议论文, 2010: 185-194

致谢

我很庆幸，能够从合肥一个小地方来到北京中科院继续我的研究生生活。我认识我的导师邹自明研究员，源于中国科大的大学生研究计划，

我第一次见到邹老师，时候已经是7月末。那个时候，正是北京最酷热的时候，少不更事的我，穿着中科大标志式的沙滩裤和拖鞋走进了导师的实验室。我看到邹老师的时候，他正在工作。我当时很是紧张，毕恭毕敬的喊：“主任好。”然后，我细细的打量着这个实验室主任来。邹老师见到我，劈头就说：“这不是中科大，请不要穿短裤和拖鞋。”听到这句话，我扑哧一下笑了，因为我知道，邹老师就是我的校友。在校友面前，我再也不用紧张了。这是我和我导师这辈子的第一条对白，但是正是这条对白，开始了我和邹老师四年的不解之缘。

永远也难以忘怀，在我充满保研未定的彷徨和不安时，邹老师直接给了我肯定的答复。而那一年，恰巧是邹老师开始他“桃李满天下”的第一年。那个时候，我固然不了解邹老师的工作是否高深，但是，我知道我应该忠于并衷心热爱我的导师。

永远也难以忘怀，邹老师在带我出去开会住宿的时候，总是把我当成自己人，直接赤裸上身，用一条大毛巾围着，就在我面前呼啦啦的晃悠，甚至咕噜咕噜的就打起呼噜了。这让我感到我的导师就好像自己亲人，我甚至觉得我的导师距离我是那样的近。

永远也难以忘怀，邹老师用以科大人固有的自尊心，来激励我勇于选择卫星轨道力学这种艰难的课程。这让我摆脱了室友和论坛上给我所带来的浮躁的心，重新找回了在科大学习时，攀越科学高峰的意气。

永远也难以忘怀，在我遇到人生价值取向的困惑时，邹老师总会把办公室的沙发当成研究生院的学堂，自己亲自给我这个学生“上课”。经常上完一次“课”，我长舒一口气，看看表竟然是连续讲了三个小时没有休息。邹老师也说过，自己好为人师，热爱传道、授业与解惑，登上“讲堂”，就宠辱皆忘了。

永远也难以忘怀，在进入研究所这种规范科研单位后，我由于深受母校的影响，保持着一种率性而行的衣着和行事风格。其实，这是不利于我与他人的合作和工作的。但是，邹老师总是竭尽自己的能力，去容忍我，包容我，保护我。这让我无比感动。

我在邹老师带领下，在从事科研项目中，慢慢的变得成熟自信起来。我一直都在用自己的努力来回报我导师对我的知遇之恩。在中科院十一五信息化的项目中，我们的工作一气呵成，毫不拖泥带水，在最终结题评比中深受专家组好评。当我回头看所有一切的时候，才深深知道，自己工作是那样的微小。而一个领导者清晰的科研思路，才是成功的关键。

我开始愈加爱戴自己的导师了。但是，在学生的心目中，邹老师是我人生的偶像，学习上的导师，是生活中的大朋友，又是我的校友和大师兄，是我孜孜不倦地追求所要成为的人。

我要深深感谢带领我一起工作的佟继周老师。佟老师是一个温柔如水的女老师。在我工作中遇到挫折时，佟老师总是用鼓励的目光告诉我，事情并没有那么糟糕。我是一个年轻人，有时候有了压力，难免有脾气，但是佟老师作为师长，总是在一旁默默的看着我，然后等我气消了，才过来温柔的安慰我。这使我感到很对不起自己的老师，但是我一直没有勇气表达这句话。我希望佟老师能够原谅学生。

我还要感谢数据网络中心的这个有凝聚力的大集体给我带来三年幸福快乐的研究生生活。感谢温和、内敛、帅气、善良的郑岩师兄在数据库设计方面给予我无私的帮助。感谢已为人父却青春洋溢的惠东师兄在项目开发的过程中，指导我解决一个又一个“拦路虎”。感谢美丽可爱的文臻师姐和文凭师姐，总是容忍我这位小师弟的“调戏”。感谢后来加入项目组的郑程师弟与我一起熬夜，最终胜利完成项目的所有工作，郑程师弟是一个优秀、纯真而朴实的人。感谢文健师妹从项目开始，就与我一起并肩作战。我在带领师妹期间，正是我研二充满锐气的时候。为此，我师妹屡屡在工作上受我的批评，我至今感到很惭愧，希望文健师妹能够原谅我这个师兄。邹老师后来批评我说：“要把身上的菱角磨圆一点，才能成为一个受人尊敬的好老师。”我至今仍在努力一点点的控制自己这种锐利得有点刺人的性格。

我要深深感谢我的爸爸妈妈。古人说：“身体发肤，受之父母。”我在北京念书期间，养成不好的生活习惯。我爸爸妈妈常常为我的健康担心，每天几乎都要给我打电话。在我人生最无助最失落的时候，是我爸爸妈妈边忍痛看着咆哮绝望的我，一边安慰我，让我重新振作起来，一步一步走向成功。而，我作为我爸妈

心中所谓的好儿子，作为我爷爷奶奶心中所谓的好孙子，作为我家乡父老眼里曾经的优秀生，选择了科研这条不归路，却未曾孝顺过爷爷奶奶和爸爸妈妈，未曾给家里带来任何经济收入，未曾尽过中国传统文化里长子应该尽的任何责任。我每每想起此事，心里就深深刺痛。

我衷心感谢在百忙之中抽出宝贵时间对我的论文进行评阅的专家们。

我要祝我自己、爱我的和我爱的人身体健康。我想这个世界再也没有比这更重大、更美丽的祝愿了。因为只有健康，才能完成实现我未竟的理想、愿望、责任和义务。

最后，愿以北宋大儒张横渠的名言共勉求学之心：“为天地立心，为生民立命，为往圣继绝学，为万世开太平！”