

天津大学硕士学位论文

虚拟天文台环境下的海量数据存储
与访问技术研究

**Research on Mass Storage and Access
Technology in Virtual Observatory**

学科专业：计算机应用技术

研究生：李文

指导教师：孙济洲 教授

天津大学计算机科学与技术学院

二零零七年六月

独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作和取得的研究成果，除了文中特别加以标注和致谢之处外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得天津大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：

签字日期：

年 月 日

学位论文版权使用授权书

本学位论文作者完全了解 天津大学 有关保留、使用学位论文的规定。特授权 天津大学 可以将学位论文的全部或部分内容编入有关数据库进行检索，并采用影印、缩印或扫描等复制手段保存、汇编以供查阅和借阅。同意学校向国家有关部门或机构送交论文的复印件和磁盘。

（保密的学位论文在解密后适用本授权说明）

学位论文作者签名：

导师签名：

签字日期： 年 月 日

签字日期： 年 月 日

中文摘要

近年来，随着观测设备和观测技术的发展，天文数据呈现爆炸式的增长，而计算机网络技术的快速发展又为这些海量天文数据的共享提供了有力的技术支持，因此，将天文学和计算机技术融合在一起的“虚拟天文台”便应运而生，使得人们突破时间和空间的限制获得丰富的天文数据资源成为可能。

然而，在虚拟天文台环境下，如何对海量数据进行有效存储成为一个亟待解决的问题；同时，如何对海量数据进行快速有效的访问，也是关系到虚拟天文台服务能力的重要课题。

本文以中国虚拟天文台为背景，研究海量天文数据的存储和访问技术。首先，对当前主要的存储技术进行综合分析和比较，结合虚拟天文台对海量数据的存储需求，设计了中国虚拟天文台的数据存储方案，并对该方案进行评价。然后，针对天文数据的存储特点，分析当前天文数据的访问现状，提出了海量天文数据的高效访问方案，对该方案进行详细论述并验证其可行性，在此基础上，提出了适应于该方案的海量天文数据分层模型。最后，在对海量天文数据的存储和高效访问研究的基础上，具体设计并实现了一套天文数据归档管理系统——FitHAS，该系统能够屏蔽不同 FITS 文件的差异，向用户提供统一的接口，并能对大量的 FITS 文件进行快速有效的访问和方便的管理。

本文在理论上具有一定创新，研究成果已经在中国虚拟天文台的建设中得到应用。实际效果证明，本文提出的存储和高效访问方案以及在此基础上实现的 FitHAS 天文数据归档管理系统可以有效满足中国虚拟天文台对于海量数据存储和访问的需求，同时具有很强的可扩展性。

关键词： 虚拟天文台 海量数据 存储 FITS SAN

ABSTRACT

In recent years, as the development of observation equipment and technique, astronomical data has showed explosive growth, and rapid development of computer network has provided strong technical support to share massive astronomical data, thus "Virtual Observatory" which integrates astronomy and computer technology emerges as the times require, which makes breaking through restriction of time and space to access a wealth of astronomical data resources possible.

However, in the Virtual Observatory environment, how to conduct mass data storage effectively has become a serious problem. In addition, how to access massive data quickly and efficiently is an important issue related to the service ability of Virtual Observatory.

This paper researches on massive astronomical data storage and efficient access technology, based on China-VO project. First, based on analysis and comparasion of popular storage techniques, in combination of massive data storage requirements of Virtual Observatory, storage scheme of China-VO is proposed, with evaluation. Then, in according to the characteristics of astronomical data storage, current existing astronomical data access technology is analyzed and effective access scheme of massive astronomical data is introduced and discussed in detail, with verification of its feasibility. On this basis, massive astronomical data layer model adapted to the scheme is introduced. Finally, on the basis of study on storage and efficient access of massive astronomical data, FitHAS, an astronomical data storage management system, is designed and implemented, which is capable of shielding differences between FITS documents and providing a uniform interface to user, as well as efficiently accessing and conveniently managing large number of FITS files.

With some theoretical innovation, the achievements of this paper has been applied in the construction of China-VO. Actual results prove that the storage and efficient accessing scheme presented in this paper, as well as the FitHAS system achieved on this basis, can effectively meet the China-VO requirement for massive data storage and access, while gaining strong scalability.

KEY WORDS: Virtual Observatory, Massive data, Storage, FITS, SAN

目 录

第一章 绪论.....	1
1.1 课题的提出	1
1.2 研究目的和意义	2
1.2.1 研究目的.....	2
1.2.2 课题意义.....	2
1.3 国内外研究现状分析	3
1.3.1 主要存储技术.....	3
1.3.2 海量数据检索与访问.....	5
1.4 本文的工作和创新	6
1.5 本文的组织	6
第二章 虚拟天文台.....	7
2.1 虚拟天文台的产生	7
2.2 虚拟天文台的研究目标	7
2.3 虚拟天文台和网格	8
2.3.1 网格.....	8
2.3.2 网格的发展为虚拟天文台提供技术支持.....	10
2.3.3 虚拟天文台是网格的试验床.....	10
2.4 China-VO.....	11
2.4.1 拓扑结构.....	11
2.4.2 体系结构.....	11
2.4.3 服务模型.....	13
第三章 海量天文数据的存储方案设计.....	15
3.1 China-VO存储现状.....	15
3.1.1 数据量.....	15
3.1.2 存储现状.....	16
3.2 China-VO对海量数据的存储需求.....	16
3.2.1 容量的可扩展性.....	16
3.2.2 可管理性.....	16
3.2.3 可用性.....	16

3.2.4 性能.....	17
3.2.5 成本.....	17
3.3 存储方案分析与设计	17
3.3.1 海量数据存储方案.....	17
3.3.2 数据备份方案设计.....	22
3.3.3 方案评价.....	22
第四章 海量天文数据访问方案的研究与设计.....	24
4.1 天文数据访问现状	24
4.2 海量天文数据访问方案设计	24
4.2.1 元数据.....	24
4.2.2 天文数据文件索引.....	25
4.2.3 典型应用.....	27
4.2.4 方案评价.....	27
4.3 海量天文数据的分层模型	28
4.3.1 物理设备层.....	29
4.3.2 数据存储层.....	29
4.3.3 元数据层.....	29
4.3.4 服务层.....	29
第五章 FitHAS天文数据归档管理系统的设计与实现.....	31
5.1 FITS文件	31
5.1.1 FITS的基本格式.....	31
5.1.2 FITS头部.....	32
5.1.3 FITS的扩展.....	35
5.2 FitHAS的设计需求.....	35
5.2.1 设计目标.....	35
5.2.2 需求说明.....	36
5.3 FitHAS的设计	36
5.3.1 概要设计	36
5.3.2 详细设计	39
5.4 FitHAS系统实现.....	44
5.4.1 包结构.....	44
5.4.2 图形界面的实现.....	44
5.4.3 异常的处理.....	45

5.5 运行结果	47
第六章 总结与展望.....	51
6.1 工作总结	51
6.2 展望	51
参考文献.....	52
发表论文和科研情况说明.....	55
致 谢.....	56

第一章 绪论

1.1 课题的提出

天文学是一门古老而又始终随着时代进步而发展的学科。作为一门以观测为基础的学科，它离不开观测数据的积累^[1]。这些资源在早期科技还不十分发达的时代，只能靠人们凭肉眼或者利用有限的观测设备去观测获取，数量上十分有限，而且在地理上的分散性也使得科学家们很难共同享有这些数据。后来随着观测设备和观测技术的发展，天文观测的数据收集能力已经有了相当大的提高，一些观测项目的数据日产出数量甚至以十亿甚至十万亿的量级来计算^[2]，所积累的数据量也正呈现爆炸式的增长，目前国际上公开的天文数据已经达到上百TB量级，不久将达到PB量级。在这种情况下，传统的数据存储和管理的方式已很难满足海量数据的存储要求，如何应对不断增长的海量数据已成为天文学界一项亟待解决的问题。

除了天文学自身的发展，计算机领域内不断涌现的新技术也使天文学有了革命性的飞跃。计算机技术和天文学相结合产生的“虚拟天文台”（Virtual Observatory，简称VO）的概念一经提出便迅速得到了世界许多国家的天文学家和计算机专家的关注。虚拟天文台是利用先进的信息技术将各种天文研究资源以统一的服务模式无缝透明地汇集到一起，形成一个统一的数据密集型的网络化天文研究与科普教育平台^[1,3]。虚拟天文台吸收了网格等先进的技术，将集成无数的资源，提供丰富的服务。在虚拟天文台环境下海量数据不但是巨大的数据资源，也是向天文学家和社会各界提供天文数据服务的重要基础。因此还需要在虚拟天文台环境下有效地组织数据资源，挖掘海量数据的潜力，从而使这宝贵的资源能为国内外的天文学家乃至社会各界所更好地使用，真正发挥其巨大的作用。

处在计算机和天文学的交叉地带的虚拟天文台，其海量天文数据的存储与访问不但是天文学家遇到的新问题，同时也是计算机技术的一个典型应用场景，是在网格技术不断发展、数据密集型和计算密集型的应用越来越多的今天所提出的一个重要课题。

1.2 研究目的和意义

1.2.1 研究目的

中国虚拟天文台(China-VO)^[4]通过多年的天文观测以及与其他国家的交流,已获得大量的天文数据,并且数据的总量还将随着各巡天项目不断的数据产出而增长。在这种情况下,本文首先要解决海量数据的存储难题,设计一种比较合理的海量天文数据的存储方案,为中国虚拟天文台建设工作的改进提供思路和建议。

其次,本课题还需要研究如何有效地归档并管理这些已有的和未来的海量天文数据,从而为China-VO的海量天文数据管理的规划和决策提供理论依据,并为进一步发布天文数据和开发数据访问服务打下良好的基础。

另外,在理论研究的基础上,本课题还实际开发具有一定通用性的天文数据的归档和管理工具,以解决目前China-VO面临的海量数据归档和管理中遇到的问题,并为未来大天区多目标光纤光谱望远镜(LAMOST)^[5]等项目的数据管理提供有效易用的工具,提高我国天文研究工作的效率和技术水平。

1.2.2 课题意义

该课题的研究在理论和应用方面都具有十分重大的意义。

首先,虚拟天文台的概念提出的时间并不长,各国的工作小组均处于研究和探索阶段,对于海量天文数据的存储的研究还不十分成熟,因此一个比较合理的存储和访问的方案将作为虚拟天文台建设中重要的一步,为进一步的科研工作打下坚实的基础,从而有力地推动天文学的发展。同时该研究成果也可为我国即将建成的LAMOST大科学项目提供一些宝贵的实践经验。

其次,该课题的研究也对计算机技术的发展起到一定的推动作用。天文学作为一个数据密集型的研究学科,为网格等技术提供了一个天然的试验床^[6],其海量数据的存储与访问的实践经验也正是对网格技术的检验和有力推动。

再次,其他一些数据密集型和计算密集型的学科如高能物理、生物技术等领域,也在走着与网格技术相结合的发展道路,也同样面临着海量数据的存储与访问问题,虚拟天文台在该方面的研究成果也将为我国这些学科的发展提供一些经验。

最后,本课题通过实际研究开发的天文数据归档和管理工具FitHAS验证了本文所论述的海量天文数据的访问方案的可行性。同时该系统也在虚拟天文台的实际应用中起到了巨大的作用,并将会为其他更多项目提供一些基础的支持。

1.3 国内外研究现状分析

近年来数据存储技术已经有了较大发展，先后出现了 RAID、DAS 以及存储网络等概念，其中一些已经比较成熟，并已在一些企事业单位中得到了运用。

1.3.1 主要存储技术

➤ RAID 技术

表 1-1 几种主要 RAID 模式比较

RAID 模式	名称	故障修复	可用容量	读性能	典型应用
RAID 0	没有容错设计的条带磁盘阵列	不支持	最大	高（盘的数量决定）	对可靠性要求不高的迅速读写。如视频、图像的编辑制作
RAID 1	镜像阵列	支持	最小	一般	需要较高数据安全保障的场合，用于小文件、随机数据写入
RAID 2	带海明码校验磁盘阵列	支持	较小	大数据量高，小数据量低	大容量数据存储，实际较少使用
RAID 3	专用校验条带阵列	支持	一般	高	需要一定可靠性的场合，适于长时间顺序访问数据的应用，如多媒体、图像处理、大型文件访问、CAD、数据仓库等
RAID 4	专用校验条带阵列	支持	一般	高	需要一定可靠性的场合，适于大文件、连续数据传输
RAID 5	分散校验条带阵列	支持	一般	高	适合用于输入/输出密集、高读/写比率的应用程序，如 ERP、大型数据处理中心的 OLTP 应用等
RAID 10	跨越镜像阵列	支持	最小	一般	要求一定数据安全性与性能场合，小文件、随机数据写入，如视频服务器系统
RAID 30	跨越专用校验阵列	支持	一般	高	要求高可用性、高速度的场合，大文件、连续数据传输，尤其是非交互的应用程序如视频流、图形和图像处理等，实际较少使用
RAID 50	跨越分散校验阵列	支持	一般	高	要求高可靠性、高速度的场合，如多事务处理、多用户存取小文件的办公应用程序

RAID (Redundant Array of Independent Disks), 即“独立冗余磁盘阵列”, 简称“磁盘阵列”, 它将多个独立的磁盘组成一个如同一块磁盘的整体, 并通过对磁盘上数据的条带化成块存取数据, 以及对一阵列中的几块磁盘同时读取, 减少了磁盘的机械寻道时间, 从而提高了数据存取速度。同时通过镜像或存储校验信息的方式, 实现了对数据的冗余保护^[7]。

RAID分为RAID 0~7 等多种模式, 还可组合成RAID 10、RAID 30 等模式, 它们具有不同的可靠性和性能, 也有各自适用的场合^[8,9]。表 1-1 中列出了几种主要RAID模式的对比^[7,10]。

➤ DAS

DAS即“直接连接存储”(Direct Attached Storage), 是早期用得比较多的一种存储方式。这种方式中将外部数据存储设备如RAID、磁带机等都直接挂载在服务器内部总线上, 这时数据存储设备是整个服务器结构的一部分, 服务器也担负着整个网络的数据存储职责。目前这种方式主要在一些小型网络或地理位置分散的网络中应用^[7]。

➤ NAS

NAS (Network Attached Storage, 网络附加存储) 方式的出现是对DAS的一个突破。它通过专用的NAS设备管理数据资源, 从而使得数据作为独立的网络节点而从服务器中分离出来。NAS的典型组成是使用TCP/IP协议的以太网文件服务器, 每个NAS拥有单独的IP地址, 可以直接挂载在主干网的交换机或其他局域网的Hub上^[7,8]。

➤ SAN

SAN (Storage Area Network, 存储区域网络) 是一种通过SCSI、SSA、ESCON和光纤通道等外围通道协议连接存储设备的后端网, 是一个高速的专用子网。SAN以数据存储为中心, 采用可伸缩的网络拓扑结构, 可以采用点对点、仲裁环、交换式架构^[11], 通过具有高传输速率的光纤通道的直接连接方式, 提供SAN内部任意节点之间的多路可选择的数据交换, 并且将数据存储管理集中在相对独立的存储区域内。SAN建立在FC (Fibre Channel, 即光纤通道) 技术之上而不是现有的IP网络基础上。光纤通道是一种数据传输技术, 其传输速率可以达到 4 Gb/s 甚至更高, 比SCSI几乎可以快上三倍。FC所采用的物理介质可以是铜线、多模光纤和单模光纤, 最远可以传送到 10km左右^[7,8,10,12]。不过SAN的成本相对较高, 相对于NAS方案来说SAN方案的成本一般会高出 20%左右^[7], 因而在一定程度上制约了其推广, 目前SAN主要还是在较大型的企业中使用。

➤ IP SAN

IP存储, 即通过IP协议进行数据交换的存储技术。它使用相对比较熟悉的IP

网络完成不同SAN之间的互连,充分保护以往基于IP网络技术的设备投资,使用户得到了比光纤通道存储局域网解决方案低得多的整体拥有成本,也使得用户可以更容易地管理SAN存储。TCP/IP网络的使用同时也打破了光纤通道目前对距离的限制。目前主要的IP存储技术包括iSCSI、FCIP、iFCP等,其中只有iSCSI获得了IETF批准,于2003年发布了其标准^[7,8,10,12],而另外两种在实践中使用较少。凭借其成本优势,目前IP SAN在市场上已开始崭露头角,对SAN造成了一定的压力。

iSCSI是一种在Internet协议网络上,特别是以太网上进行数据块(block)传输的标准,是一个供硬件设备使用的可以在IP协议上层运行的SCSI指令集。利用iSCSI,用户能够在诸如高速千兆以太网上进行路由选择,从而通过现有的TCP/IP网络来构建存储局域网,同时也解决了SCSI的访问和距离问题^[7,8,10,12]。

1.3.2 海量数据检索与访问

在虚拟天文台环境下,海量数据的访问要求传统的数据访问应用或服务能够应对前所未有的海量数据的压力。天文领域的传统服务和应用多数是需要访问并处理文件的,在面对海量文件时,只能依次对每个文件都进行类似的解析和处理,工作量将是十分巨大的,在响应速度上很难达到用户的要求。而且有很多情况下,并不是所有文件都需要被处理,但天文数据文件本身的特性使得应用程序往往要读入文件之后才可以判断出该文件是否为所需文件,因而大大降低了工作效率。

为了在海量数据中迅速定位和访问到目标文件,需要有海量文件检索技术的支持。在这方面,在文档库中检索关键词的全文检索技术目前已经取得了较大进展。该技术包括两种实现方式:一是根据检索表达式直接在原文档中匹配查找,二是对文档预先建立索引,在检索时再对索引进行检索。后者可以使用正排表及倒排表组织全文的索引,这方面的应用前景非常广阔,因而很多学者也对此进行了大量的研究工作^[13-15]。然而这类检索方法一般是检索哪些文件中出现指定关键词以及判断关键词出现的频率,可以说是一种针对文本的相似程度的判断和评价方法。而天文学数据的处理有很多是通过将文件中的属性信息取出并进行一定的运算和处理从而得出结果的,比如指定观测目标所在的天区,然后根据各数据文件中的坐标信息计算该图像是否属于该天区,从而进行精确的筛选。这与上述关键词的检索还是有一定区别的,因而难以直接应用上述成果,而应该研究更适合天文数据特点的海量数据检索的技术。

1.4 本文的工作和创新

本文根据中国虚拟天文台的数据存储现状、特点以及自身的发展定位，讨论了目前 China-VO 天文数据的存储中存在的问题，并提出一种更有效的海量天文数据存储方案，作为进一步的数据访问服务的基础。

本文分析了目前 China-VO 的数据访问方式在面对海量数据时所存在的不足，提出并讨论了一种解决方案，然后通过实例验证了其可行性和有效性。在此基础上又进一步提出一种虚拟天文台环境下海量天文数据的分层模型并予以阐述。

此外，本文还在理论分析的基础上，针对天文观测中广泛使用的 FITS 文件的数据访问需求，实际开发了 FitHAS 系统，作为 China-VO 海量数据存储和访问管理的一个重要组成部分，有效地完成天文 FITS 文件的归档和管理。

1.5 本文的组织

第一章概述课题的提出背景、研究目的和意义，以及国内外研究现状等。

第二章介绍虚拟天文台的相关背景知识。

第三章首先分析了中国虚拟天文台的数据存储现状存在的问题，然后结合 China-VO 的海量数据存储需求，在综合比较几种数据存储技术的基础上，提出一种海量天文数据的存储方案，并进行了详细论述和评价。

第四章根据目前中国虚拟天文台的数据访问方式和状况，提出一种更有效的数据访问方案，并结合实例研究该方案的可行性及有效性。在此基础上进一步提出并论述了一种海量天文数据的分层模型。

第五章首先介绍了天文 FITS 文件的格式标准，然后结合 China-VO 对天文数据管理的实际需求，详细阐述了通用的 FITS 文件归档管理系统——FitHAS 系统的设计与实现。

第六章对本文工作进行总结，并就下一步的工作做出展望。

第二章 虚拟天文台

虚拟天文台是古老的天文学与最新的信息技术相结合的产物，它利用最先进的计算机和网络技术将各种天文研究资源，包括各种观测数据、天文文献、计算资源、存储资源、各种软件工具甚至天文观测设备，以某种统一的服务模式无缝地汇集在系统中。它可以为天文学家提供丰富的资源和强大的服务，使他们从数据收集、数据处理这些繁琐的事务中彻底摆脱出来，从而将精力集中于自己感兴趣的科学问题上。虚拟天文台给天文学研究带来了一个革命性的发展，必将成为开创“天文学发现新时代”的关键性因素^[16]。

2.1 虚拟天文台的产生

二十世纪九十年代，探测器、计算机和互联网领域所取得的突破性进展使得天文数据资源急剧增长。已经实施和计划中的巡天计划产出的星表、光谱、图像等数据已达TB量级，并逐渐逼近PB量级^[17]。如此巨大的数据产出，在天文学历史上第一次使天文学家得到的数据多得用不了。数亿个天体目标多个波段的数据资料的产生对数据处理、分析工具提出了更高甚至全新的要求，也为数据挖掘、高级模式识别、大规模交叉相关统计分析等这些新兴科学方法提供了用武之地。此外也在天文学历史上第一次使得高度复杂数值模拟结果可以与观测数据进行有科学意义的相互验证，结束了数值模拟结果因缺乏足够的观测数据而很难验证的历史。巨大而复杂的观测数据不断产生，迫切需要得力的工具、方法来发现蕴含在这些数据海洋中的宝藏^[18]。

为此，1999年，美国科学院天文与天体物理发展规划委员会在名为“新千年的天文学和天体物理学”的未来十年发展规划中把建立（美国）国家虚拟天文台（NVO）作为最优先推荐的项目之一，首次提出了虚拟天文台的概念。此后，世界各国的天文学界迅速响应，纷纷提出了各自的虚拟天文台研究计划，在全球掀起了一场虚拟天文台的研究浪潮^[16]。

2.2 虚拟天文台的研究目标

虚拟天文台提供了各种资源，例如数据资源、计算资源、存储资源，并将各

种资源无缝透明地融合在一起,向用户提供可以轻松地进行在线研究的门户和窗口。不同的虚拟天文台项目提出了不同的科学目标,主要包括^[19]:

(1)系统地探索数据

通过对多波段多尺度多时段的天文数据的探索,可以研究宇宙的大尺度结构和银河系结构、AGN 物理和演化、低表面亮度星系,并且可以探索以前很少研究的观测参数空间。

(2)时间域和图像域的探索

研究不同时间域的数据,可以发现一些变源或变化的天文现象。对于图像数据,可以将源的形态和环境有效地参数化,通过多尺度分析或借助模式识别方法发现新的现象,如引力透镜的发现。每一种方法都有自己的偏差和极限,可根据数据的特点和科学要求来选择合适的方法。

(3)稀有的或未知类型的天体或新现象的发现

天文学家最关注的事情,就是希望发现一些稀有的或未知类型的天体(如高红移类星体、II 型类星体、T 矮星等)或新现象,从而推动理论的不断发展和完善。

(4)巨型的数值模拟与观测数据的对比

以前由于存储和计算资源的缺乏而不能实现的任务,现在借助虚拟天文台则可以实现。

(5)多波段天文学、统计天文学和交叉学科的发展

多波段数据在虚拟天文台中的有机融合为多波段天文学研究提供了很好的平台,也促进了交叉学科的发展。

2.3 虚拟天文台和网格

2.3.1 网格

Ian Foster在 1998 年出版的《网格: 21 世纪信息技术基础设施的蓝图》一书中将网格描述为“构筑在互联网上的一组新兴技术,它将高速互联网、高性能计算机、大型数据库、传感器、远程设备等融为一体,为科技人员和普通老百姓提供更多的资源、功能和交互性。……让人们透明地使用计算、存储等其他资源。”2000 年, Ian Foster进一步将网格描述为“在动态变化的多个虚拟机构间共享资源和协同解决问题”。2002 年 7 月,他又限定网格必须同时满足三个条件: (1) 在非集中控制的环境中协同使用资源; (2) 使用标准的、开放的和通用的协议和接口; (3) 提供非平凡的服务。至此, Ian Foster已经给出了狭义的网格的概念^[20,21]

还有许多赞同广义的网格概念,即巨大全球网格 GGG(Great Global Grid),它不仅包括计算网格、数据网格、信息网格、知识网格、商业网格,还包括一些已有的网络计算模式,例如对等计算 P2P(Peer to Peer)、寄生计算等。

不管是狭义还是广义的网格,其目的不外乎是要利用互联网把分散在不同地理位置的电脑组织成一台“虚拟的超级计算机”,实现互联网上所有资源的全面连通^[20]。

网格系统大致可以分为五个基本层次:构造层、连接层、资源层、汇集层和应用层^[6],如图 2-1 所示。

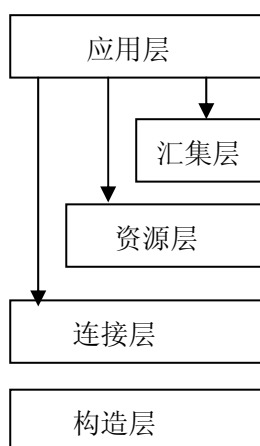


图 2-1 网格体系结构

构造层:本地控制系统的接口,提供了共享的资源。比如计算资源、存储资源、数据、网络资源和传感器。构造层部件执行的是本地的、与资源相关的操作。

连接层:安全方便的通信,定义了网格环境下的网络交易所需的核心通讯和认证协议。通讯协议实现了构造层资源间的数据交换。认证协议建立在通讯服务基础上提供用于用户和资源识别的加密的安全机制。

资源层:共享单一资源,为个体资源的共享操作定义了安全对话、初始化、监测、控制、记帐、付帐等协议。资源层协议针对的是单个的资源,不考虑分布式资源集合广域上的状态和管理等事务。

汇集层:协调多样资源。资源层关注的是单个的资源,汇集层则关注的是资源的总体。建立在资源层和连接层构成的协议瓶颈上,汇集层实现了广泛的共享操作。汇集层的功能是屏蔽网格资源层中各种资源的分布、异构特性,向网格应用层提供透明、一致的使用接口。汇集层也称为网格操作系统,它同时需要提供用户编程接口和相应的环境,以支持网格应用的开发。

应用层:是用户需求的具体体现。在网格操作系统的支持下,网格用户可以

使用其提供的工具或环境开发各种应用系统。

网格技术在很大程度上与虚拟天文台的开发目标一致,利用网格技术作为基础设施发展虚拟天文台是必然趋势。一方面,虚拟天文台的建立和实现需要网格技术的支持。另一方面,虚拟天文台将为网格技术的发展提供最好的试验床^[6]。

2.3.2 网格的发展为虚拟天文台提供技术支持

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

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

2.3.3 虚拟天文台是网格的试验床

天文数据有着其他学科数据所无法比拟的特点:

- ◆ 天文数据绝大部分是开放数据。国际上许多大型天文观测项目的观测数据都会及时在互联网上公布,这为数据共享提供了良好基础。
- ◆ 天文数据很少有商业价值。这意味着相互之间不存在保密性的问题,可以将数据和结果进行自由共享,非常适合国际性的联合研究与试验。
- ◆ 天文数据是海量的数据。例如美国正计划建造的“暗物质望远镜”,每天观测数据高达 18TB。这非常适合网格环境下海量数据处理和数据挖掘算法的研究。
- ◆ 天文数据是真实而归档完好的数据,并提供在线访问服务。
- ◆ 天文数据是高度复杂的数据。海量高度复杂的数据对数据处理、数据挖掘、可视化等研究提出了挑战。

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

2.4 China-VO

“中国虚拟天文台（China-VO）”计划是以中国科学院国家天文台为首的中国天文界在 2002 年提出的。中国虚拟天文台将作为未来国际虚拟天文台的有机组成部分，为中国天文界提供一个全新的网络化研究平台，引领中国天文学进入数据密集型在线科学研究新时代。

2.4.1 拓扑结构

图 2-2 是China-VO的拓扑结构图^[6]。国家天文台（NAOC）、国家天文台云南天文台（YNAO）、中科大天体物理中心（USTC）、北京大学天文系（PKU）、清华天体物理中心等天文单位将构成China-VO的应用主体。

国家网格（CNGrid）拥有强大的计算资源和 IT 技术优势，比如上海超级计算中心（SSC）、中科院计算机网络信息中心超算中心（SCCAS），中科院计算机网络信息中心科学数据网格（CNIC），清华大学高性能计算研究所（THU）等，将成为 China-VO 的大规模计算平台和技术试验床。China-VO 将利用中国国家网格（CNGrid）这一良好桥梁，与国内 IT 领域的兄弟单位合作，抓住“网格”这一新的互联网发展机遇，提高我国的网格研发和应用水平。

作为 IVO 不可或缺的一部分，China-VO 将与全球的 VO 项目一道为实现 IVO 的宏伟目标共同努力。

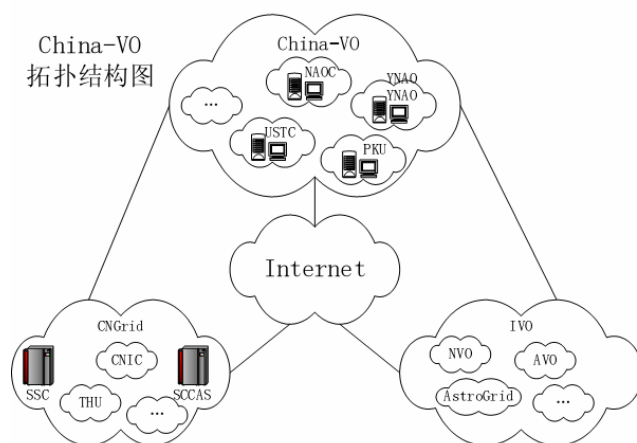


图 2-2 China-VO 拓扑结构

2.4.2 体系结构

依据网格的分层结构以及实际的发展定位，China-VO 设计了自己的体系结

构，从基本的功能模块角度可以表示为图 2-3 所示的架构^[6]。

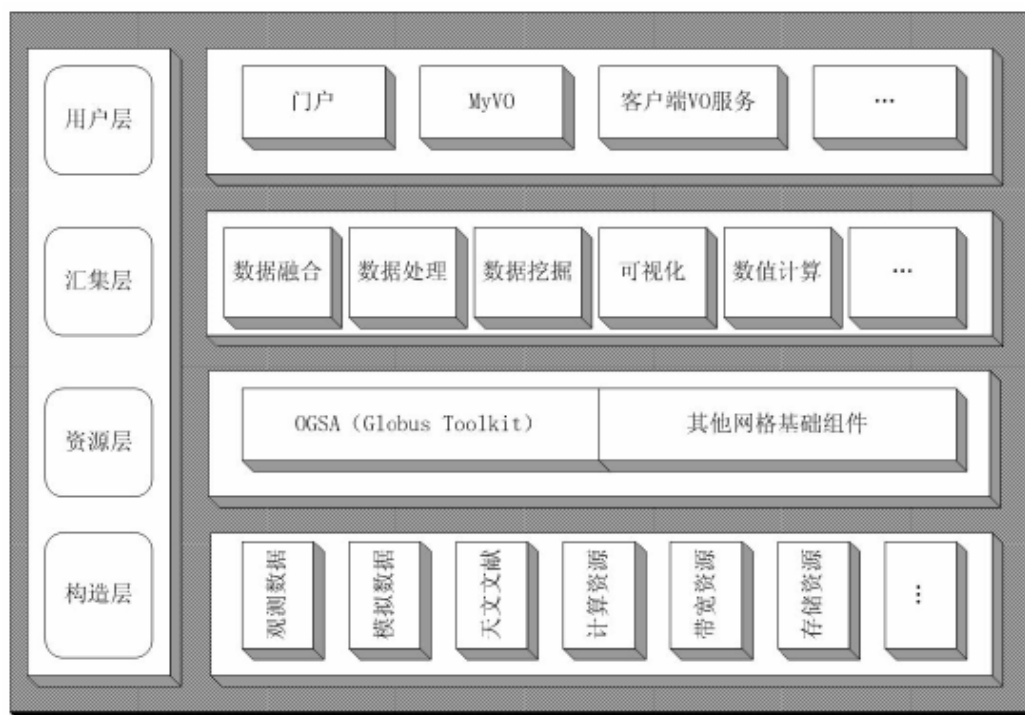


图 2-3 China-VO 体系结构

整个体系结构分为四层，从下到上依次是构造层、资源层、汇集层和用户层。

构造层是中国虚拟天文台系统的资源基础，其中包括各种天文数据资源、计算资源、网络资源、存储资源等。各种天文数据资源在虚拟天文台这样一个数据密集型在线研究平台中占有非常关键的作用，是 VO 成功运作的基础和前提。

资源层将以开放网格服务架构为基础，配合其他网格系统工具，利用标准的数据模型和服务模型，通过抽象化实现统一的资源访问，特别是统一的数据访问和统一的计算访问，以及网格系统管理等功能。这里，系统管理主要涉及作业管理、安全管理、资源状态管理、数据管理等。

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

用户层，包括 VO 客户端服务和 VO 门户，是整个体系的最高层，直接与虚拟天文台用户接触。用户层的基本职能是用户任务提交和处理结果返回，主要功能包括用户登录、身份认证、VO 资源浏览、任务编制和提交、结果显示、数据下载、偏好设置，以及运行于用户本地的 VO 客户端应用程序等。

2.4.3 服务模型

在 China-VO 的研发过程中采用了网格技术中面向服务的设计理念。

面向服务的模型有许多优点，环境中所有部件都进行了虚拟化，通过层层抽象以统一的方式对待这些服务。所有服务都要提供一系列标准接口，在此基础上，可以构造层次式的高级服务。虚拟化还能实现多个逻辑资源实例向同一物理资源的映射，在无需考虑底层资源的实现和管理情况下构造上层服务。服务的虚拟化还提供了共同语义行为向本地主机环境的无缝映射，从而实现平台无关性^[3]。

China-VO 是一个数据密集型的在线研究平台。为了实现数据密集型在线研究这个目标，它必须实现三个方面的基本功能：数据访问、数据处理、数据互操作。这三个基本任务在 VO 系统中可由三个角色来承担：数据服务提供者(Data Service Provider, DSP)，应用服务提供者(Application Service Provider, ASP)和注册(Registry)。把 VO 系统的基本功能服务按照所处角色不同将其分配到 DSP、ASP 和 Registry 中，便得到 China-VO 的服务模型，如图 2-4 所示^[6]。

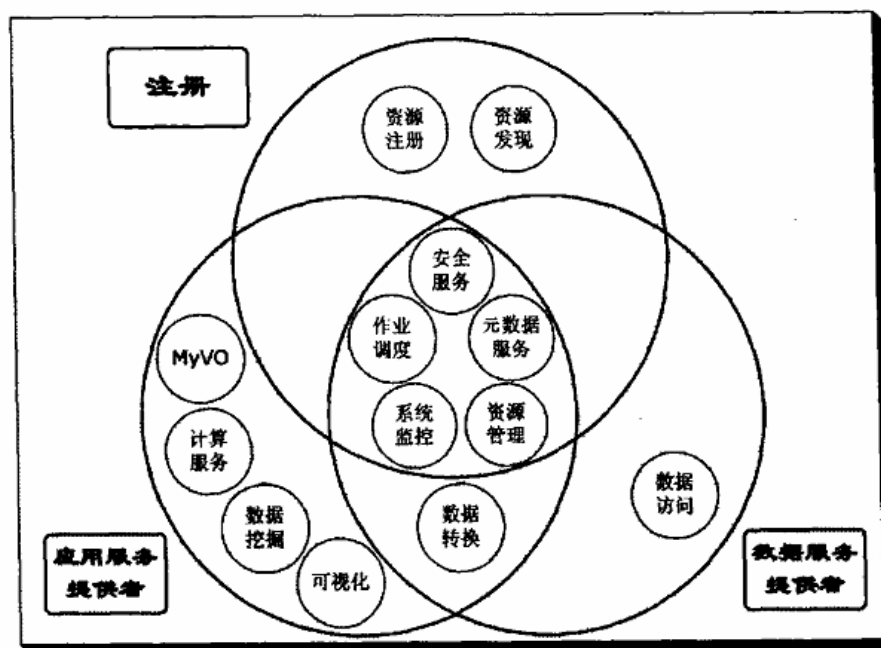


图 2-4 China-VO 服务模型

从模型中可以看出，资源管理、系统监控、安全服务、任务调度对所有角色都是必须的，将由符合 OGSA 标准的网格操作系统来提供。虽然元数据服务对所有角色也是必须的，但由于这部分涉及许多天文学的学科特点和内容，所以 China-VO 将参考 OGSA 体系提供的元数据服务机制给出 China-VO 的元数据服务。资源注册与发现、数据访问、计算服务、数据挖掘、可视化、MyVO 等服务

则将由 VO 社区独立开发或者借鉴其他学科领域的实践经验来实现。

第三章 海量天文数据的存储方案设计

虚拟天文台作为数据密集型的天文研究平台，丰富的数据资源、高效的资源管理和强大的信息处理能力是其内在的要求。在当前观测数据猛增的时代，如何合理有效地解决海量数据资源的存储问题，从而为进一步的数据访问和信息处理提供保证，是虚拟天文台建设过程中必须要面对的首要问题之一。

3.1 China-VO 存储现状

3.1.1 数据量

目前，我国的LAMOST、太阳空间望远镜（简称SST）、五百米口径射电望远镜（简称FAST）等一批大观测设备在投入使用后，均将以巨大的数据产出极大地丰富我国的天文数据资源^[17,22]。其中LAMOST项目在全部投入使用后，将可能达到每晚积累 20GB 原始观测数据的速度。如果按照每年观测 300 天，则一年积累原始数据 6TB。在 5 年的一个周期中，至少获得 30TB 的原始数据量。若加上其他数据则整个项目的存储需求将可能达到 60TB^[23]。

除了这些自产的天文数据，国家天文台天文数据中心还积累了大量的国际天文数据，其中包括SDSS等著名巡天项目的观测数据。SDSS一期巡天 5 年来已经积累了大量数据资料，仅DR5 这一个版本总数据量就达近 15TB，其数据量情况如表 3-1 所示^[24]。

表 3-1 SDSS DR5 的数据资源量

图像数据	图像	9.0 TB
	星表 (DAS, fits 格式)	1.8 TB
	星表 (CAS, SQL 数据库)	3.6 TB
光谱数据	2D 光谱	60 GB
	一维光谱	170 GB

China-VO 不但需要存储来自这些观测项目的已有的观测数据，还要应对不断增长的数据，这都是对 China-VO 数据存储能力的巨大挑战。

3.1.2 存储现状

China-VO的数据存储基本上走的是“Linux + IDE硬盘 + 软RAID技术”的道路。从2001年ADS镜像站点的550GB服务器,到2002年天文数据中心的4.5TB数据服务器,虽然存储容量有所增加,但基本还没有摆脱传统的存储方式^[25]。目前主要还是通过将磁盘挂接在服务器上实现数据的存储,并且简单地靠增加磁盘完成容量的扩充。这种方式虽然在短期内看起来比较容易实现,但不利于长期的发展。首先是该方式容易产生信息孤岛,数据分散在各服务器上,不利于统一管理,而且有可能造成存储空间的浪费,对于各磁盘上的剩余空间难以有效地联合到一起进行统一规划和使用。其次是存储设备的扩展受到了很大的限制,系统的扩展性较差。另外,由于存储任务需要由服务器来担当,因此也极大地影响了服务器的性能。可见,China-VO目前的存储方式很难满足未来海量数据的存储需求,需要推行更合理的存储方式以适应虚拟天文台的发展。

3.2 China-VO 对海量数据的存储需求

3.2.1 容量的可扩展性

虚拟天文台对观测数据存储的要求中,首要的就是对存储容量的要求。没有足够的存储空间放置数据,那么观测仪器得到的数据只能丢弃,进一步的研究工作也将因此无法开展。所需要的存储容量不仅要能容下当前国内观测和国际共享得到的所有观测数据的总量之和,而且还要具备足够的可扩展性,以便应对数据量持续增长的状况。

3.2.2 可管理性

数据存储的建设不但需要初期建设的投入,而且也需要做好长期的管理工作。一个好的数据存储设计应该使存储设备和数据更易于管理,可以提供比较全面的管理功能,并且有友好易用的管理工具,从而有效降低存储管理人员的培训费用和工作难度。

3.2.3 可用性

天文观测数据是进行科学研究的宝贵资料,由于该学科的特殊性,很多原始观测数据的观测条件很难重现,数据一旦丢失或损坏将造成极大的损失。因此,使用的存储设备和整个存储方案需要有较高的可用性。

3.2.4 性能

由于虚拟天文台环境中,不但有以科研机构为主体所参与的计算密集型和数据密集型的科研工作,还将同时向公众提供多种数据服务,因此虚拟天文台应能在有大量并行请求的情况下仍保持一定的性能。

3.2.5 成本

在满足其他几项需求之外,还应合理地控制成本,这不仅包括建设初期的投入,还需考虑后期扩容时的成本,因此,目标是要寻求一种总拥有成本(TCO)较低的方案。

3.3 存储方案分析与设计

3.3.1 海量数据存储方案

根据上述对虚拟天文台的存储需求和现状的分析可知,我国虚拟天文台的数据存储宜采用比较成熟的技术和标准,一方面可以降低建设工作的难度、成本和风险,另一方面可以使整个存储系统具有一定的通用性和良好的可扩展性,从而使上层的数据服务和应用开发更易于管理和维护。

天文原始观测文件等数据的存储一旦发生问题将会造成相当大的损失,为尽量避免这样的损失发生,不但要考虑购置具备足够容量的设备,还可以采用RAID和存储网络等解决方案,以提供一定的数据冗余及高级的存储安全保障^[26]。

(1) 磁盘阵列模式分析

对于 China-VO 目前已有的磁盘可以通过 RAID 技术组合起来。通过对目前常用的几种磁盘阵列模式的比较,不难发现各模式均有各自不同的适用场合。在海量数据存储方案的设计中,并不一定要选功能最强大的,也不是选最便宜的,而是选择最合适的模式。

根据 China-VO 数据存储的可用性的要求,应该选则具有一定冗余或校验策略的模式。RAID 0 由于不提供冗余支持,无法满足数据存储的可用性要求,因而首先被排除。

未来虚拟天文台的很多数据访问服务将会向天文学家和社会各界开放,这就需要数据存储方式达到大量并发性请求对吞吐量和延迟的要求。另外绝大多数操作是对数据的读取,写入操作相对很少。从这一点来看,所选的 RAID 模式应该具备较高的读数据的性能以及对大量随机数据请求的支持。而 RAID 3 和 RAID

30 因为使用并行访问方式,更适合大型数据的顺序访问,因此也不符合 China-VO 的要求。

从容量的可扩展性角度考虑,由于 RAID 4 模式使用单一校验盘,该校验盘也可能成为容量扩展时的系统瓶颈,当磁盘数目增多时将尤为明显,因此 RAID 4 并不适合虚拟天文台这种对容量扩展性要求很高的场合。

在余下的几种模式中,很难找到 RAID 2 的实际商业产品^[10];再考虑成本的因素,只有 RAID 5 才更符合 China-VO 的要求。。

综上所述,RAID 5 是可以满足 China-VO 存储需求的一种较适合的模式。它可以使用软件实现也可以使用硬件搭建。一般而言,通过软件实现 RAID 的方法在整体性能上不如硬件实现的 RAID,因此在财力允许的情况下,China-VO 可考虑购置相应硬件设备搭建 RAID,以进一步提升性能,在当前使用服务器挂接存储设备的状况下有效降低服务器的负荷。另外,这也是将来升级到存储区域网络的一步有效措施。

(2) 网络存储的设计

位于北京的国家天文台是 China-VO 中一个主要机构,担负着 China-VO 中很大一部分的数据存储和管理的工作。因此 China-VO 海量数据的存储方案设计可以从两个层面进行,一方面要对国家天文台的存储方案进行分析和设计,从而为 China-VO 亟待解决的数据存储问题寻求一种可能的途径,同时还要进一步从更高的角度探讨 China-VO 整体的存储问题,以提高数据存储未来的扩展能力。

目前包括国家天文台在内的很多机构均采用由服务器负担存储任务的方式,这实际也是一种 DAS 存储方式。这种方式的缺点在于,服务器需要承担数据存储的任务,因而极大地影响了性能。同时随着更多的存储设备的加入,将导致存储孤岛数量增多,为管理带来了很大的难题。

SAN、NAS、IP SAN 将存储从应用服务器中分离出来的特性,打破了 DAS 对扩展存储容量的局限,并且有利于存储的管理,比 DAS 更适合未来的存储发展。因此网络存储将是海量天文数据存储的首选方式。

从架构上考虑,这三种技术有着较大的差别。NAS 可以较充分地利用原有的 TCP/IP 网络,不过受传统以太网的单段距离限制^[7]; SAN 需要构建光纤通道替代原有的 TCP/IP 网络连接,可以突破 NAS 的距离限制而达到 10km 的长度; IP SAN 通过传统的 IP 技术连接多个 SAN,覆盖范围又可以远远大于 10km,所以一般是用在将距离较远的几个 SAN 网络进行互联,在短距离内反而体现不出优势。目前位于北京的国家天文台,其存储和应用服务设施位置比较集中,并未出现这种需要联结多个 SAN 网络的需求,因此对于国家天文台的存储建设,IP SAN 还不能充

分发挥出其优势。当然，在未来China-VO可以集合国内各天文台和其他高校甚至天文学家自发组织的研究小组，地域范围将会更广，这时就需要类似IP SAN这样的方式将各子机构进行互联。

在扩展性方面，SAN所采用的网络结构允许用户自由增加磁盘阵列、磁带库等存储设备，并且这些设备一般可以被自动识别，因而具有较高的扩展性。NAS也具有一定的扩展性，但需要配置专用文件服务器，后期扩容时还将面临成本问题。IP SAN借助TCP/IP网络可以将更多的SAN网络连接起来，依靠各SAN内部的容量扩展以及所连接SAN子网的数目来实现容量上的扩展。在WAN上，iSCSI凭借自身的寻址和发现机制可以拥有甚至比光纤通道技术更好的扩展性^[27]。因而SAN和IP SAN在扩展性方面都是China-VO的可选方式。

在性能方面，SAN在三者中是最高的。NAS由于需要使用网络进行备份和恢复，因此占用了一定的网络带宽，使得其存储性能难以与SAN相比。iSCSI由于其潜在的丢包和慢启动的问题使得性能也难以与FC SAN匹敌^[27]。无论是考虑数据的传输速度还是整个存储网络对并发请求的吞吐量、响应时间，SAN都是一个可以满足China-VO包括国家天文台的需求的选择。

从可用性角度，NAS专用的文件服务器可能存在单点故障，因而可靠度并不是很高。SAN允许进行镜像配置，从而可以改善存储系统的容错能力^[7]。而IP SAN因架构在传统的TCP/IP网络之上，其可用性很大程度上要受TCP/IP网络的影响。因此，相比之下，SAN是其中可用性相对较高的一种。

对于成本，SAN在建设和管理维护方面都是三者中最高的。不过NAS的成本优势在后期扩容时将逐渐淡化。而IP SAN的优势主要体现在使用已有TCP/IP网络替代光纤通道的那一部分，以及原有网络产品的利用方面。由于RAID可以纳入到SAN网络中，所以原有网络产品的投资还可以得到一定保护；而对于国家天文台而言，由于地理分布相对集中，可以借用原有网络替代光纤通道的部分所占比例很小，成本上的优势也并不十分明显。

另外，IP SAN中唯一的一个已推出标准的iSCSI协议，虽然已在市场上出现了一些产品，但是市场认可度还是比较低，所占市场份额仍较小，相应产品线也不如SAN和NAS的完备，标准也尚需时日加以检验和完善，短期内不宜直接运用到China-VO的建设中。

通过上述比较和分析，可以看出，SAN成本最高，不过在性能、可扩展性、可用性方面表现都十分突出。NAS具有一定的成本优势，但是其他方面都不很理想，难以满足China-VO的要求。同样具有成本优势的IP SAN在连接距离方面也颇具优势，但在性能、可用性方面均略逊于FC SAN。

◆ 国家天文台数据存储方案

对于国家天文台而言，性能、可扩展性、可用性等均是优先考虑的因素，这样 IP SAN 因为性能上的劣势并不适合用于其数据存储，而同时其距离优势和成本优势也未能充分体现，没有必要为节省成本而使用这种尚不十分成熟的技术。因此 SAN 将是国家天文台的一个非常适合的选择。

对于 SAN 目前在成本上的劣势，考虑到国家天文台的发展现状和建设水平，不能强求将存储建设得一步到位，可以采用逐步升级的策略。国家天文台目前所购置的大量磁盘暂时采用 DAS 和软 RAID 的方式提供存储和访问服务，在进一步扩充和升级时可以增加具备 SAN 组网特点的存储设备，并且组建硬 RAID 以提高性能；在条件基本成熟之后，可以升级到 SAN 存储方案，将已有的存储设备纳入到 SAN 网络中，以保护先期投资。这样既实现了海量数据存储设施的建设，又充分考虑到了初期建设资金比较匮乏的现状，是种比较可行的方案。

由此，可以得出国家天文台海量数据存储的一种设计方案如图 3-1 所示。

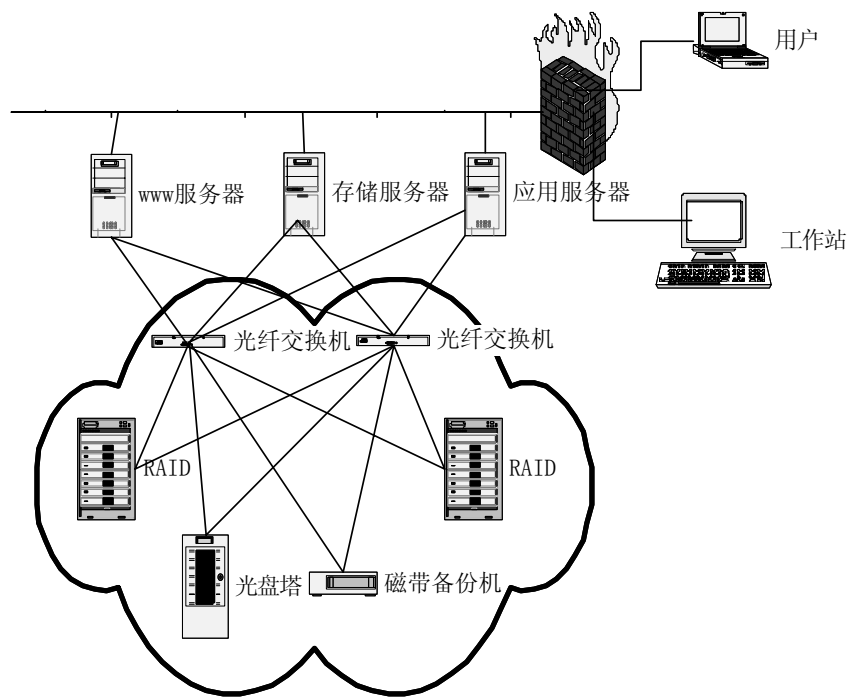


图 3-1 国家天文台的 SAN 存储结构图

在本方案中，原有的磁盘阵列等存储设备被纳入 SAN 网络，实现集中式的管理。各服务器均可以通过光纤交换机与存储设备相连，各存储设备也可以通过任一条光纤交换机与存储网络外的服务器相通。在服务器部分，将不同的应用分布在不同的服务器上，同样提高了可用性。可以配备专用的服务器，进行存储网络的数据备份、恢复等数据维护和管理工作。整个存储网络和服务器部分都可以

隐藏在防火墙之后从而得到一定的安全保障，用户通过网络来访问 China-VO 对外提供的服务。

◆ China-VO 的整体数据存储方案

在更广域范围内考察，China-VO 具有国家天文台所远不及的覆盖面积，包含多个这样 SAN 或者其他的存储架构。这时 SAN 就难以发挥作用了。

这种情况下，IP SAN 恰恰可以发挥其在距离以及成本上的优势，将相距较远的各机构连接起来。利用 iSCSI 等技术，将在各 SAN 中的 SCSI 命令等封装成 IP 包在传统网络上进行传输。虽然在性能上较依赖 TCP/IP 网络的表现，但是这却是在成本、性能、可靠性等因素之间权衡的结果。由此得到的设计方案如图 3-2 所示。

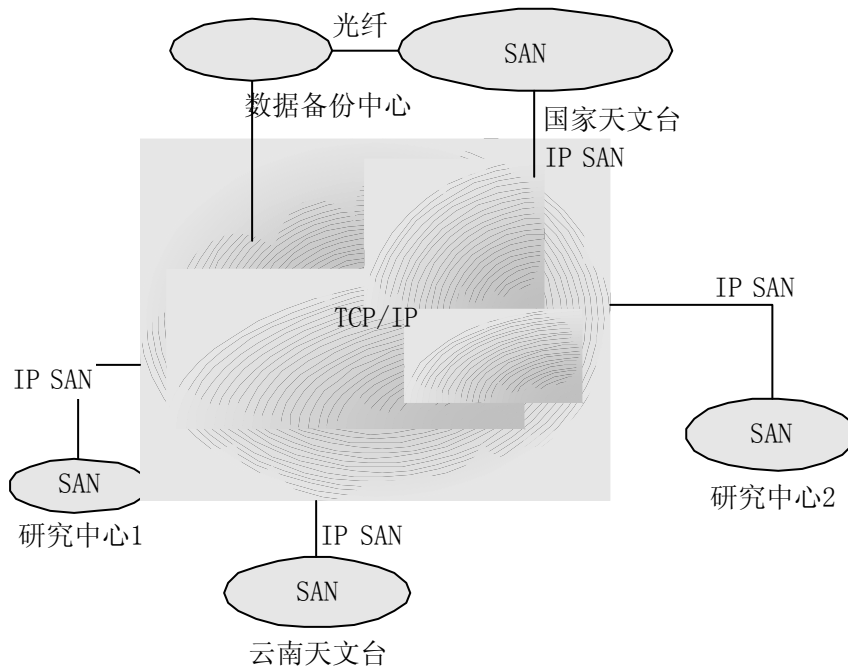


图 3-2 China-VO 存储方案示意图

由于 IP SAN 技术目前还不十分完善，因而 China-VO 的存储建设可以作为国家天文台建设之后的下一步工作进行。可以预见的是，高性能的 SAN 和成熟的 IP 技术的结合是种必然趋势，IP SAN 技术必将发展得更加成熟，而其定位也十分符合 China-VO 发展的定位，因此 China-VO 选择 IP SAN 技术应该说是可以适应未来技术发展的。

3.3.2 数据备份方案设计

天文数据的珍贵要求 China-VO 必须做好数据的备份和恢复工作。对于从国外获得的数据集镜像, 比如 SDSS 等, 由于数据量庞大, 不适宜在数据发生损坏后再次从国外传输数据, 而应在国内做好数据的备份工作。

数据的备份可以先在同城进行。例如可以借助中科院科学数据库等提供的存储服务完成海量数据的同城备份^[28]。由于距离相对较短, 同城备份可以采用光纤等介质互联, 以获得更高的性能, 数据恢复的时间相对也会更短一些。但是同城备份的缺点在于发生类似地震一类的灾害时, 无法保证数据的安全可靠。

对数据安全性要求较高的一些行业和单位, 如金融、保险等, 常采用三地容灾的方式保护他们的数据。对于 China-VO 而言, 可以在同城备份之外, 在远程的另一数据中心(如云南天文台)建立一个完整的数据镜像。在进一步的存储建设中, 可以将远程的数据中心也升级为 SAN 网络存储, 利用上述联结各数据中心的 IP SAN 架构, 可以容易地实现异地数据镜像复制, 同时也可相互间的访问提供更好的连接基础。

3.3.3 方案评价

上述存储方案不但要求 China-VO 在现有存储方式的基础上购置新的设备, 而且在拓扑结构上也有一定的变化。对于该方案的可行性和有效性, 可以根据 China-VO 对数据存储的各方面要求进行评价。

- ◆ 容量的扩展性: 在 SAN 网络中添加新设备或摘除旧设备都是比较容易的, 许多光纤网可以自动辨别网络拓扑结构, 使 SAN 的扩展更加容易。而光纤通道光纤网使用的是 24 位地址结构, 最多可容纳 1600 万个设备^[12]。即使按照每个阵列 1 TB 的容量计算, 16000 PB 的总量也足以应当虚拟天文台的容量要求。
- ◆ 可管理性: 相对于目前的 DAS 方式而言, SAN 实现了存储设备的集中管理, 因而比 DAS 可管理性更高。相比使用传统 TCP/IP 网络的 NAS 和 IP SAN 方式, 对存储管理人员的要求就显得要高一些。但是 SAN 将是一种发展趋势, 在对管理人员进行相关培训上的开支还是值得的, 况且目前 IBM 等公司已经开发出比较成熟的 SAN 管理软件^[12], 可以很大程度上降低管理人员的工作难度。
- ◆ 可用性: 与 NAS 相比, SAN 对数据备份和恢复功能有更好的支持, 因而提供了更高的可用性。同时, 设计方案中使用了两个光纤交换机, 有效地避免了单点故障问题, 使系统可用性进一步提高。假设使用一个光

纤交换机的可靠度能达到 99%，则在使用两个光纤交换机时，仅当两个都发生故障时才使整个存储网络瘫痪，即这时可靠度将达到 $1-(1\% \times 1\%) = 99.99\%$ 。

- ◆ 性能：SAN 所采用的光纤通道技术提供了较高的性能，目前主流的速率已达到 4 Gbps，而千兆以太网 1 Gbps 的速率也只有 SAN 的四分之一，因此性能比基于 TCP/IP 的 NAS 和 IP SAN 都要更高一些。同时 SAN 有效地将数据的备份和恢复的传输请求局限在存储子网中，避免了对 LAN 的性能的影响，从而可以有效满足虚拟天文台复杂、大量、并发的数据访问服务的需求。
- ◆ 成本：通过采用逐步升级的方式，可以有效减少初期的资金投入，在后期扩容和升级时，也可以有效保护投资，很好地满足了用户对成本的要求。

综上所述，该方案有效地满足了 China-VO 对海量数据的存储需求，是一种比较合理而且可行的方案。

第四章 海量天文数据访问方案的研究与设计

在国内许多行业中，都普遍存在着数据信息利用不充分的现象，大量的数据仅以物理状态被简单地存储着，同时由于无法实现海量数据的高效搜索，使得这些数据无法为企业的发展提供决策支持^[29]。在天文学领域同样面临这样的问题。拥有海量的天文数据并不是虚拟天文台的最终目的，真正的目的是如何利用好这些数据，更好地促进科研工作的开展，并推动天文知识的普及。因此，虚拟天文台的这些海量天文数据不但要有合理、可靠的存储作为坚强的后盾，更要有高效的数据访问机制为国内的研究人员和天文爱好者提供优质的数据服务。

4.1 天文数据访问现状

天文数据包括星表、图像、光谱等多种类型^[30]，另外，一些天文方面的文献和观测日志也作为天文数据的一部分，同样需要得到很好的存储和管理^[3]。这些数据一般以文件或数据库形式保存。在天文学中，比较常见的数据文件格式标准有普通文本文件、FITS、VOTable^[31]等，目前很多天文观测项目生成的观测数据就是以FITS文件形式保存的，而很多数据查询服务将结果包装成VOTable形式返回给用户^[2,32]。FITS格式的文件一般用于记录天文图像数据，很多虚拟天文台研究小组已开发出针对FITS文件进行解析和可视化处理的工具。这些大多是直接处理文件而且是少量文件的应用。

对于海量数据的使用，高级用户可能是从PB级的数据中进行一定的文件筛选，然后再将得到的文件（仍可能是GB或TB级）进行进一步的科学计算、数据挖掘等处理，这也是数据密集和计算密集型的工作。初级和中级用户则一般是从海量的数据中进行相对简单的查询。根据这种访问需求可以对China-VO海量数据的访问方案展开讨论。

4.2 海量天文数据访问方案设计

4.2.1 元数据

在天文学领域，像FITS和VOTable这样的格式标准中往往具有一定的结构化

特征，比如FITS文件中会出现以ASCII编码的用于描述图像位置信息等的部分，这些部分是“描述数据的数据”，因而又称为“元数据”。在不同的学科，“元数据”也有着不同的具体含义。在天文学领域，“元数据”是用于描述档案、档案提供的服务、其中的数据集合、每个数据集合的结构和语义（即含义）以及数据集合中每个数据集的结构和语义。典型的元数据可以是用来描述星表、图像或光谱等数据对象的信息。例如，FITS文件中所记录的天文图像的语义元数据就是这幅图像FITS头部的逻辑内容^[30]。

天文数据文件的元数据包含了一些重要描述信息，而其中很多信息正是天文学家在检索和筛选数据文件时所需要的判断依据，如坐标信息等；同时还有一些可以作为简单数据处理所使用的输入信息。因而元数据也在一定程度上“代表”了该文件的一些重要信息。

4.2.2 天文数据文件索引

在虚拟天文台的海量数据访问方案中，海量数据的检索是其中最重要的内容之一，解决这一问题无论对初、中级用户还是高级用户的需求，都可有效地提升海量数据访问服务的水平。

由于天文数据具有文件写入和归档之后几乎很少修改的特点，可以预先为海量天文数据建立索引，在文件检索时直接对索引进行操作，从而有效完成海量文件的检索工作，同时也不必担心数据文件更新引起索引频繁变动的情况发生。

接下来就要解决几个问题：一是从文件中抽取什么信息用以建立索引，二是怎样抽取这些信息，三是如何建立索引。

◆ 文件索引

使用文件索引的一个显著优点就是可以在相对较少的信息中快速找到所需信息，并能方便地定位到原文件，这在全文检索技术中已有了较好的体现^[13]。在海量的天文数据文件中建立索引同样也要达到这样的目的。因而首先要选取可以“代表”文件的信息，上面所述的元数据就可以达到这一要求。但是仅是文件中提取的信息还不够，为达到“定位文件”的要求，必然还需要在元数据中增加关于文件的定位信息，在windows、linux等操作系统中，这也就是文件的路径信息。拥有这部分信息，就可以向用户提供检索所需要的条件字段，并且文件的定位信息也将在得到检索结果后方便用户直接去访问原文件。

◆ 索引信息的提取

由于许多天文数据文件的元数据具有特定的格式，因而可以根据其格式规范，预先对所有文件进行扫描，依次获取每个文件中的元数据，同时记录下文件

的定位信息。这样就得到了这许多文件的索引信息的集合。

◆ 索引的建立

一般情况下,文件的索引可以保存在文件系统中,也可以存入数据库,甚至还可以存在系统内存中。由于虚拟天文台有着海量的文件,因此将大量信息放在系统内存是不可行的,这会严重影响系统的性能。因而需要在文件系统和数据库管理系统之间进行比较。

如果将这些元数据信息保存在文件系统中,则在进行海量文件访问和检索时,需要编写程序对预先提取的元数据信息执行查询,以完成类似“该文件中的数据是否在指定天区范围内”这样的计算和处理过程,如果针对文件系统进行优化则性能会更好一些。而若存入数据库,则是利用数据库管理系统中的 SQL 语句执行这样的查询工作。数据库同样支持查询的优化,一些数据库管理系统绕过了文件系统而实现自己的 I/O 操作,因而性能方面也不会太差。那么对于这两种方式的取舍可以从扩展性以及已有应用的兼容性等方面考虑。

SQL语句可以实现丰富灵活的查询,支持对列的操作和统计,以及更复杂的存储过程等特性。另外,数据库管理系统还可以对数据挖掘提供更好的支持。在 China-VO 的实践中,已实现了很多基于数据库的服务,同时也已开发出了 VO Query Language (VOQL) 来扩展数据库的查询功能,并对不同数据库的差异进行屏蔽。因此,使用数据库保存文件索引可以具备较高的可扩展性并有效降低维护成本,同时也有利于已有软件的集成^[6]。

另一方面,国际上许多虚拟天文台小组也倾向于选择数据库来作为他们服务后端的数据资源,而虚拟天文台的技术发展也将可能允许从单一的访问点访问这些大量的数据库从而增大天文研究的可能性^[33]。

基于以上考虑,China-VO 可以选择 DBMS 存储数据文件的元数据作为文件索引。当然不同 DBMS 有着很多差别,比如作为商业产品的 Oracle 数据库表现就比较优秀,而免费数据库中,MySQL 和 PostgreSQL 的功能上并没有本质的差别^[6],实际应用中,可以根据具体需求选择合适的数据库。

在性能上,像这样在数据库中存储数据文件元数据的方案,将原来直接访问文件,变成了先检索数据库再访问文件,而且多占用了一定的存储空间。但是综合起来看,通过存储元数据建立文件的索引信息,有效地实现了海量数据文件的检索,同时可以避免潜在的大量文件 I/O 操作、文件解析处理以及对压缩文件的解压缩处理等操作,而且数据库 SQL 语句的强大功能和对优化的良好支持也都提供了较大的可扩展性,整体还是具有一定优势的。可以说,这种方案更像是一种用空间换时间的策略。

4.2.3 典型应用

为进一步验证上述存储元数据方案在虚拟天文台中的可行性,本文以一个典型的 FITS 文件访问的场景为例进行讨论。

FITS, 即Flexible Image Transport System, 是世界各天文台之间用于数据传输、交换的统一标准格式。在FITS的一般结构中, 含有若干使用ASCII码编写的描述信息, 用于描述数据数组的结构、坐标系统以及附加参数等^[34,35]。比如, 其中的RA、DEC等参数提供了一些图像的位置信息, 那么就可以通过一些复杂的坐标变换获得这个FITS文件中图像所覆盖的天空区域。

在一个典型的场景中, 一位天文学家可能需要处理一些 FITS 文件, 也许他只想处理一小部分天区的数据, 但是当有数以万计甚至更多的文件摆在他面前, 而且每个文件所覆盖的天区信息需要通过读取文件中头部的记录获得, 这时, 他就需要逐个读取这些文件, 并根据其中的位置信息判断这是不是他所需要的文件。即使不考虑针对每个文件计算其覆盖天区的复杂程度, 单是对文件的遍历以及大量频繁的 I/O 操作便会使他的工作效率大大地降低。

如果采用了上文所述为文件建立索引的方式, 通过使用数据库存储文件元数据, 包括 FITS 文件头部记录中的各个属性、参数所提供的信息, 如图像的位置信息、观测设备、观测时间等, 以及文件资源的定位信息, 那么虚拟天文台不但向外界提供了更方便的数据库查询接口, 同时也不失定位原文件的功能。对于普通用户来说, 可以像使用其他数据库应用一样进行天文数据的查询, 查询过程交由数据库来完成; 对像这位天文学家这样的高级用户, 还可以在检索之后, 进一步获取对应文件, 进行后续的研究工作。

4.2.4 方案评价

由上面的科学范例不难看出, 通过存储文件元数据建立文件索引的方案有效提高了工作效率, 是在提供海量天文数据的访问过程中一种比较有效的策略。该策略具有以下特点:

- ◆ 必要性: 单纯依靠遍历访问如此大量的数据文件的做法效率是比较低的, 使用更高效的策略是一种必然趋势。
- ◆ 可行性: 对于这个典型的场景而言, 因为 FITS 有标准的格式约定, 并且其头部记录部分采用 ASCII 格式编写, 因此可以方便地提取, 也十分适于在数据库中进行灵活的查询。由于 FITS 在天文学领域的广泛应用, 也使得在更大范围应用该策略成为可能。
- ◆ 经济性: FITS 文件的大小从几 K Byte 到几十 M Byte 都有可能, 以天文

台已归档的部分 SDSS 的光谱数据为例, 某一目录下的 FITS 文件平均大小都在 160K 左右, 640 个文件总大小约为 100M, 而存储入数据库的元数据, 实际占用的物理存储空间不过几十 K。如此推算的话, 1T 的这样的 FITS 文件, 存入数据库的元数据所占用的空间约有几百 M, 还是可以接受的。因此以相对较小的空间代价换取更好而且更重要的时间性能, 可以说是比较经济的解决方案。

- ◆ 易用性: 数据库中强大的 SQL 语句和数据库所支持的一些查询优化方法, 使得用户可以比较容易地达到在海量数据的检索结果和检索时间上的要求。
- ◆ 关联性: 在入库的元数据中, 还可以扩充其他重要信息, 例如文件的位置, 这样就构成了元数据和对应文件之间的一一映射, 应用程序便可据此定位到文件从而进行进一步的操作。如此, 便不仅支持更高效的检索, 而且还可以方便地访问文件。

目前 China-VO 针对 BATC 项目的需求, 已选择了 MySQL 数据库进行数据文件元数据的存储, 并实现了该数据集的发布, 实践证明该方案收到了良好的效果。

4.3 海量天文数据的分层模型

通过上面的分析可知, 为获得更高的海量数据文件的访问效率, 可以将文件的元数据存入数据库, 对数据文件的索引信息进行统一管理和维护, 也就是在数据文件的层次之上构建一个可供应用程序直接访问的“索引层”或者“元数据层”。由此可以将 China-VO 的海量文件的存储抽象为一个 4 层的模型, 如图 4-1 所示。

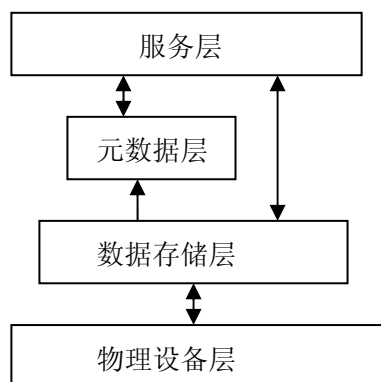


图 4-1 海量天文数据的分层模型

4.3.1 物理设备层

这一层主要包括物理存储设备，主要研究如何连接、管理这些存储设备，是上面各层的基础。

该层可以有多种不同的实现，前文提出的 China-VO 的存储方案就可以作为其中一种实现方案。使用 SAN 存储结构可以很好地满足存储容量、可扩展性、可靠性和高性能等要求，同时又能有效地实施管理和进行升级，并且即使在设备故障或升级过程中，也不会对上层起到很大的影响，较好地达到分层模型中的屏蔽底层具体实现的目的。

4.3.2 数据存储层

这一层主要是实际存储于基础设施上的数据，或者说相当于是一些比特流。这一层存储的实体可以以文件的形式存在，如包含天文图像信息的 FITS 文件等。

从数量上说，这一层是十分庞大的，储存着海量的数据。这一层的数据可以直接被一些应用程序或者服务所访问，如可以解析并处理 FITS 文件的软件等；也可以通过上面的“元数据层”而与“服务层”联系。该层的数据如果不能被很好地使用和处理，则只能是一堆无用的比特流。只有有丰富的服务访问这些数据或者可以有效地从中提取出元数据加以利用，才可以最大地发挥这些海量数据的作用。

4.3.3 元数据层

这层主要是从“数据存储层”的数据文件中抽取的元数据。按照基于计算机科学的规范去描述，这一层存储的相当于是比数据抽象程度更高的信息，一般存放在数据库中以元数据的形式进行管理^[36]。通过这一层的桥梁作用，可以使基于海量数据的数据访问服务获得较大的性能提升。

元数据层显然要比数据存储层要小一些，因为从数据文件中的“元数据”抽取出来的是一个更小的数据集合，这里的“更小”包括两个维度：

- ◆ 数据集中的元素个数，也是需要将元数据存储入库的数据文件个数。在数据文件都参与元数据提取的情况下，这个维度上的大小并没有变化。
- ◆ 元数据的属性数目。只将元数据所涵盖的属性映射到数据库中，而不是文件中的全部内容，也相当于减少了对应的数据库的字段数目。

4.3.4 服务层

在网格环境下，很重要的一个概念就是服务。很多数据处理功能将以服务的

形式进行封装供用户使用，而在“数据存储层”的数据也正是通过该层的服务得以与外部的用户实现共享。

这一层可以提供大量的服务，因此这里的分层模型也具有与网格沙漏模型相类似的形状，即具有较大的数据存储层和服务层，而中间的元数据层却较小。

第五章 FitHAS 天文数据归档管理系统的设计与实现

在海量天文数据的分层模型中,元数据层的存在一定程度上提高了海量数据的访问效率,这种方案的关键在于从数据存储层中获取元数据来构造元数据层。

目前 China-VO 也正面临着这样的问题,即一方面拥有海量的观测数据文件,另一方面,一些数据访问服务需要读取的是数据库中的元数据记录而不是原始数据文件本身,由此产生的数据“断层”现象现在还缺乏有效的工具予以消除。因此 China-VO 迫切地需要这样一个工具。在这种情况下,一个通用的 FITS 文件元数据归档和管理系统——FitHAS 天文数据归档管理系统便应运而生。

5.1 FITS 文件

天文学灵活图像传输系统FITS (Flexible Image Transport System) 最早是由美国D.C.Wells等人于 1979 年提出的。1982 年国际天文协会(IAU)接受并推荐将其作为天文台之间进行图像数据交换的标准格式。FITS描述了数据定义和数据编码的一般方法^[37],对一维、二维、多维的数据类型都提供了合适的转换,不仅适用于天文数据,还可用于其他学科的图像处理^[34]。

在实践中FITS不但加入了对更多数据结构的支持,也推广到了更为复杂图像格式的传输,不论是综合孔径、甚大阵、CCD观测记录,还是CDS发表的天文星表和CD-ROM星表,都提供了较好的支持。如今,以FITS格式保存的天文观测数据已占有相当大的比例,并且数据量还在以较快的速度增长^[34,35,38]。

5.1.1 FITS 的基本格式

一个 FITS 文件的一般结构包含一系列逻辑记录 (logical record), 逻辑记录的长度总是 23040 bit 即 2880 字节,因此一个 FITS 文件的大小是 2880 的整数倍。

在FITS文件的这些逻辑记录中,一般会包含一个或多个“头部数据单元”(Header and Data Unit, 简写作HDU), 每个HDU是由一个“头部”(header, 又称“标题”、“文件头”)以及其所描述的“数据数组”(data array, 又称“数据矩阵”)所组成, HDU也可以只包含头部而没有数据数组。在一个FITS文件含有多个HDU时, 第一个头部数据单元也就是主头部数据单元 (primary HDU), 包含文件主头部 (primary header), 以及主数据数组 (primary data array)。简单的FITS

文件可以只包含主头部和一个主数据数组，这就构成了FITS基本格式（basic FITS）^[34,39,40]。

表示图像数据的FITS文件中，图像数据以FORTRAN-77 规则存储，低字节在前，因此在某些计算机上（DEC PDP_11）需要对二进制数据进行“字节交换”的操作以使其符合FITS标准。若图像数据不满 2880 的倍数，则用 0 填满。当然图像数据也可以为空。在头部使用ASCII编码记录图像的其他属性，描述数据数组的结构、坐标系统以及传送任一附加的参数和伴随的正文。随着随机组扩展（random-groups extension）、ASCII表扩展等规则的加入，一个FITS文件在主头部数据单元之外，具有了存储和表达更多信息的能力^[34]。

5.1.2 FITS 头部

FITS的头部是通过一组卡片映像（Card image）来描述数据数组的。每一个卡片映像都是一个 80 字节长的使用ASCII编码的字符串，作为一个逻辑实体，可以描述数据数组中的一项信息。这些卡片映像的最后需要以“END”关键词作为结束标志，其后直到头部结束都用空格填满^[34,41]。一个典型的FITS头部可能具有如图 5-1 所示的结构。

```

SIMPLE =                T / Standard FITS file
BITPIX =                8 / No. of bits per pixel
NAXIS =                2 / No. of axes in matrix
NAXIS1 =               123 / No. of pixels in X
NAXIS2 =               321 / No. of pixels in Y
EXTEND =                T / There may be FITS extensions
BLOCKED =              T / The file may be blocked

CDELTA1 =              -2.3 / Coordinate increment
CRVAL1 =              -73.3 / Coordinate of reference pixel
CRPIX1 =               12.0 / Reference pixel in X
CTYPE1 = 'PIXEL      ' / Coordinate type of X axis

CDELTA2 =               7.1 / Coordinate increment
CRVAL2 =              300.1 / Coordinate of reference pixel
CRPIX2 =              -11.0 / Reference pixel in Y
CTYPE2 = 'PIXEL      ' / Coordinate type of Y axis

OBJECT = 'Ramp 8-bit'  / Name of image
ORIGIN = 'ESO          ' / File was prepared at ESO-Garching
DATE   = '19/08/92'   / Creation data of this file

COMMENT This test file was created by P.Grosbol, ESO (pgrosbol@eso.org)

COMMENT Simple 8-bit ramp pattern for testing of FITS readers

END

```

图 5-1 FITS 头部

(1) 卡片映像

在 FITS 的基本格式中，头部是由若干行信息（又称“头部记录”或“标题

记录”)组成,每行记录描述了后面图像数据的一个参数值,如数据数组的结构和坐标系统等信息。一个 2880 字节的逻辑记录中刚好含有 36 个这样的头部记录。头部记录使用一个长 80 字符的卡片映象 (card image) 来表示,这些卡片映象具有同样的语法格式,如下:

关键词 = 参数值 / 注释说明

每个卡片映象中的关键词 (keyword) 是用 ASCII 码编辑的 8 字符的字符串,可以由大写英文字母、数字、下划线 “_” 和连接符 “-” 组成。左对齐放在 1—8 列,若不满 8 字符,则右边用空格填充。如果给关键词进行了赋值,在第 9 列为 “=”,第 10 列为空格,所赋值可以是数值、字符和逻辑值。

- ◆ 逻辑值: (T 或 F) 在第 30 列
- ◆ 整数值: 右对齐在 11-30 列里,若有虚部右对齐在 31-50 列里
- ◆ 浮点值: 要求 10 进制小数点,如果用指数计数法,则右对齐在 11-30 列里
- ◆ 字符值: 在一对单引号中,第一个引号在第 11 列,字符串的长不应少于 8 字符 (不包括空格),不足时在后面同样使用空格填充

注释 (Comment) 在数值后,一般用 “/” 开始,区分大小写。虽然不是必需的,但一般建议使用注释,以使得数据信息更清晰易读。

(2) 必需关键字

在 FITS 文件中,对于图像数据的描述,有些关键词是必需的,因此可以定义一组最小设置的关键词,如表 5-1 所示。

表 5-1 FITS 头部中的必需关键字

关键词	类型	含义
SIMPLE	逻辑	值为 T 表示文件符合基本 FITS 标准,否则为 F
BITPIX	整数	指明用于表示每一个象元值的位数
NAXIS	整数	图象里坐标轴数
NAXISn	整数	当 NAXIS 大于 0 时,其后必须按次序跟有 NAXIS1、NAXIS2 等等,表示每维的大小。如 NAXIS1 指明在数组内沿最快变化轴的象元数,NAXIS2 指明在数组内沿次快变化轴的象元数
END	无	最后的卡片映象

(3) 任选关键词

对于天文图像数据来说，坐标信息的辅助参数对明确解释数字图像十分重要，有一些任选的关键词可以用于描述图像的强度、坐标和文本信息，这些也是推荐使用的。表 5-2 列出了一些在天文图像中常见的任选关键词。

表 5-2 FITS 头部中的任选关键词

关键词	类型	含义
BSCALE, BZERO	浮点数	将磁带上的象元值转换为真实值的参数。 实际值=BZERO+BSCALE×图像值
BUNIT	字符	强度单位
BLANK	整数	如果某象元未定义值，则磁带上该象元对应位置上写入该数值
OBJECT	字符	图像名，通常是观测对象的名字
DATE	字符	写 FITS 文件的日期（‘日/月/年’）
DATE_OBS	字符	数据的观测日期（‘日/月/年’）
ORIGIN	字符	写磁带的研究单位
INSTRUME	字符	数据的观测仪器
TELESCOP	字符	所使用的望远镜
OBSERVER	字符	观测者姓名
CRVALn, CRPIXn, CDELTn, CROTAn	浮点数	定义第 n 个轴的象元坐标与真实的物理坐标之间的关系
CTYPEn	字符	定义第 n 个轴的物理坐标类型
DATAMAX, DATAMIN	浮点数	给出文件中数据的最大和最小值
EPOCH	浮点数	坐标系统的历元（以年为单位）
空白, COMMENT, HISTORY	无	注释

除此表中的关键词外，还有其他一些可选的关键词可以使用。FITS 格式并未对卡片映象的数目和种类进行限定，所以还可以有更多的关键词扩展进来。

5.1.3 FITS 的扩展

基本 FITS 格式的定义为其他类型的数据提供了足够的灵活性，可以较好地支持 FITS 格式的扩展。

在 FITS 的一般扩展中，允许一个单一的 FITS 文件里包含一系列头部/数据组，每一组的开始都有一个头部，且均与文件主头部具有同样的一般格式。在作为扩展的头部/数据组中，头部的第一个卡片映像描述了随后的扩展数据数组的类型，因此解析 FITS 文件的程序可以跳过一个未知的扩展类型，到下一个头部/数据组^[34]。

5.2 FitHAS 的设计需求

5.2.1 设计目标

在中国虚拟天文台所获得的数据之中，有大量的观测数据是以 FITS 文件形式存储的，如何将这数据更有效地组织起来从而支持更快、更灵活的数据服务是课题组需要研究的问题之一。同时，存在于 FITS 格式本身的文件之间的差异性，也增加了为外界提供统一的数据访问接口的难度，如何屏蔽底层的格式差异，也是有效共享这些数据的一个重要课题。

基于此，中国虚拟天文台迫切地需要这样一个工具，在已具有巨大的数据量而且总量还在不断快速增长的情况下，完成海量数据文件的更快速有效的组织和元数据的入库归档，同时又能高效地屏蔽数据文件之间的格式差异，使之有效地统一于虚拟天文台这样一个数据资源共享的平台之中。

FitHAS，即 FITS Header Archiving System，就是针对 China-VO 的这个现状而产生的，具有一定通用性的海量天文数据的归档和管理系统。它的设计目标是：

- ◆ 通用性：智能地识别符合 FITS 规范的各种格式的 FITS 文件，有效地屏蔽已有数据文件之间的格式差异，并最大限度地适应未来 FITS 的格式扩展。
- ◆ 高效性：支持海量数据文件的统一管理，通过简单的设置和有限的操作完成一整套归档管理的流程，并且可以支持大量文件成批处理。
- ◆ 易用性：为用户提供更友好的接口，除一定的交互功能之外，还要有效地将内部的工作流程进行“封装”，为用户提供比较全面的服务，而不需要他们具备关于文件格式或者文件存储形式的基础。
- ◆ 容错性：有效兼容不同格式的 FITS 文件，整个数据归档管理的操作过程不会受到其中任何因格式错误或其他情况无法正常处理的文件的影响，

例外的文件将会被推迟操作并生成详细的结果报告，以进行进一步的处理。

5.2.2 需求说明

根据虚拟天文台的数据现状和对数据访问服务的整体规划，FitHAS 应具有如下功能：

- FITS 文件头部的解析和浏览
- FITS 文件元数据和数据库中表模式的自动映射
- 存储 FITS 元数据的数据库表的模式的自动生成
- 单一 FITS 文件的入库归档
- FITS 文件批量入库归档
- 数据库中元数据表的浏览
- 提供数据管理的可视化界面和完整归档管理流程的用户向导
- 归档文件的过滤器筛选：只有文件名与指定模式相匹配的文件才可以进行进一步的解析或者入库归档。
- 元数据记录的冲突检测与处理：若某条元数据与已存在于表中的记录是重复的记录，则提供 3 种处理策略：
 - 跳过冲突的记录，将其他记录照常导入；
 - 按照从文件中新解析得到的信息更新元数据表中的相应记录，其他无冲突的记录照常导入；
 - 将元数据表清空，再导入记录。这种策略一般是用于全面更新当前元数据表，以保证表与数据文件的同步和一致性。
- 国际化支持
- 文件归档参数设置

5.3 FitHAS 的设计

5.3.1 概要设计

根据实际的需求，将 FitHAS 设计为平台无关的独立启动的应用程序，可访问本地/异地的文件和数据库，并通过可视化界面与用户进行交互。

FitHAS 的体系结构主要分为核心事务层和用户接口层，如图 5-2 所示：

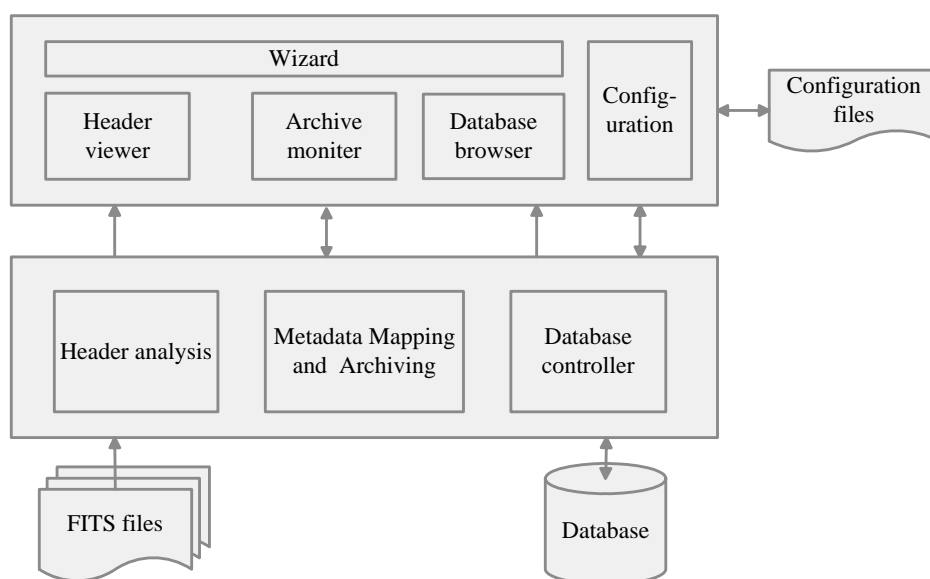


图 5-2 FitHAS 体系结构图

系统的核心事务层从用户接口层接收一些必要的参数,通过对文件系统和数据库操作系统的访问,并执行相关的数据处理逻辑,从而进行数据归档和管理的核心操作。最终还将生成操作报告传递给用户接口层。

用户接口层主要负责与用户的交互。除了提供可视化的视图和命令行方式的控制台,还将从用户处接收参数并将操作报告反馈给用户。系统的参数设置在用户接口层提供交互接口,允许配置文件的读取和更新。

按照用户的使用偏好,在用户界面层可将各功能按照数据归档向导的逻辑流程进行组织,将它们按照一定的顺序划归于 4 个归档向导之中,因此在这一层将提供如下几个模块:

- ◆ FITS 头部浏览
- ◆ 数据库元数据表的浏览
- ◆ 单一 FITS 文件的归档处理向导
- ◆ FITS 文件的批量归档管理向导

各模块的工作流程分别如图 5-4, 图 5-5, 图 5-6, 图 5-7 所示。

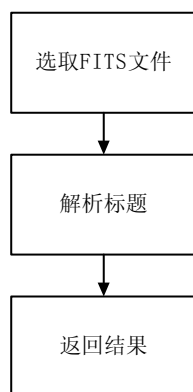


图 5-4 FITS 头部浏览向导的流程

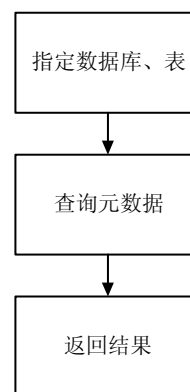


图 5-5 数据库元数据表浏览向导流程

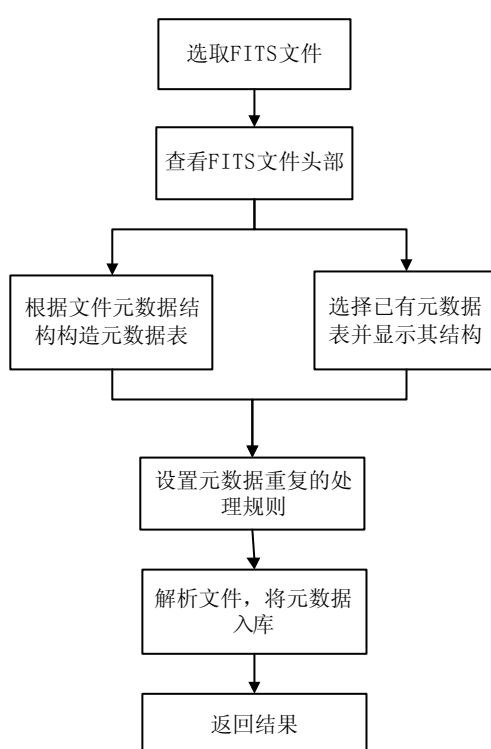


图 5-6 FITS 文件归档向导流程

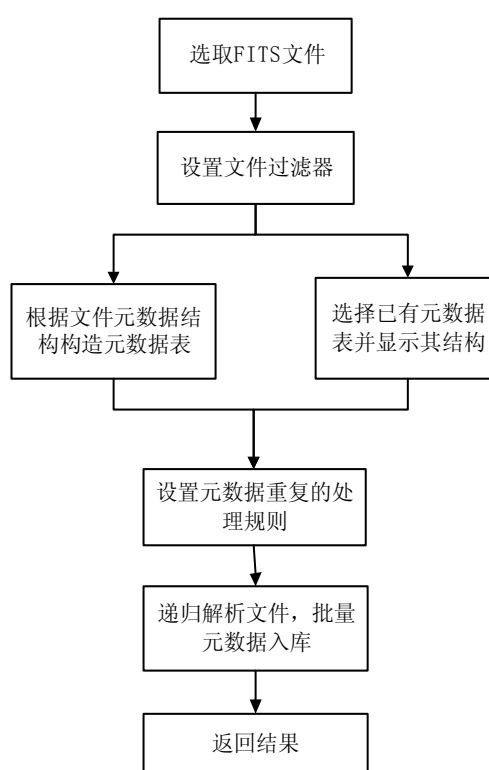


图 5-7 FITS 文件批量归档向导流程

这些向导主要使用了几个由核心逻辑层提供的功能模块，分别是：FITS 头部解析、FITS 元数据与表结构的映射、元数据入库。针对这几个核心模块的处理逻辑，首先需要约定相关的处理规则。

5.3.2 详细设计

(1) 元数据映射规则

FITS 头部的每行卡片映象信息都可以使用类似<K, V, C>的三元组的结构表示, 其中 K 表示关键词, V 表示值, C 表示注释。因此 FITS 头部也就可以表示为一些三元组的集合, 即{<K, V, C>}。

根据卡片映象的结构, 可以将每个卡片映象的信息映射为一个字段, 整个 FITS 文件头部将被映射为数据库中的一条记录, 即:

- 文件 <-> 数据库记录
- 数据描述属性 <-> 字段

一般可以在一张表中, 只映射同一种来源、同一种类型的 FITS 文件, 如 SDSS 的 DR4 的光谱数据, 这样各文件的头部结构大体相近, 元数据更能方便准确地反映原始数据的状况, 易于管理。

然后通过使用如下规则, 将元数据映射到数据库中:

- 关键词 <-> 字段名
- 取值 <-> 字段取值
- 注释 <-> 字段的说明

◆ 关键词和值的映射

FITS 格式对头部结构的约定比较明确, 一般地, 可以解析头部每一行卡片映象的前 8 个字符, 去除空格得到的字符串部分即是关键词。可以将关键词直接作为元数据表的字段名称, 这样后续的导入元数据的工作也会非常方便。在映射的过程中, 需要注意避免使用数据库的保留字作为表名或字段名, 比如关键词“DEC”在 MySQL 中是个保留字, 不可作为字段名, 但是利用 MySQL 的单引号和反单引号的特殊意义, 就可以继续使用这个单词, 方法是使用反单引号“`”将保留字括起来即“`DEC”来构造这个表的字段。

◆ 类型映射

FITS 格式约定了整数、浮点数、逻辑值、字符等类型, 然而不同的数据库所支持的字段类型有较大差异, 也均与 FITS 格式的定义有差别。因此需要定义一种数据类型的映射和转换规则, 并且对于不同的数据库设计不同的转换规则。

通过将元数据表的字段定义为字符串类型 (如在 MySQL 中的“varchar”、“text”等), 可以使数据表中的元数据几乎完整地保持其在文件中的原貌。除了 FITS 文件中字符类型的关键词的值将比文件中少了左右单引号之外, 其他类型均保持原状, 而且使用指数计数法表示的浮点数类型的关键词也仍以原来的形式

表示，不会造成精度的损失。使用字符串类型保存元数据是一个可以保证容错性和完整性的映射原则，对于不同数据库，应依据此原则有针对性地选用相应类型去构造元数据表结构，比如在 MySQL 中，字符串类型包括“varchar”、“text”等，由于 MySQL 对于 varchar 型字段的总的长度有上限的要求，因此这里选用了 text 类型。其他数据库的情况不再赘述。

◆ 字段匹配

在文件中元数据与数据表的结构不匹配时，如果是元数据表的当前字段未在文件头部中出现，则此字段留空；而如果是文件头部中的关键词在元数据表中没有相应字段，则这条记录不能被导入数据库，这可以通过进一步地修改元数据表的结构或者选择一个匹配的元数据表之后再次尝试导入，以达到完整保留 FITS 元数据的目的。

◆ 注释处理

FITS 格式的灵活性不但表现为允许出现新的关键词，而且在某些格式的约定上也会出现差异，最常见的是注释的用法。这里的“注释”主要是指整句或者整段的说明，是与其他普通关键词一样可以整行出现提供完整信息的一类关键词，不包括在卡片映象格式中所定义的针对某个关键词的解释内容。比如，以下几种都是出现过的注释：

a) 标准型：

标准型注释的格式如图 5-8 所示，前 8 个字符是关键词的位置，第 9 列是等号，第 10 列为空格，后面使用单引号将注释内容括起来，因此注释的内容应应为两个单引号之间的字符。

```
SKY      = 'best      ' / best= latest reduction, target= the i
COMMENT  = 'Target is useful to see why we did what we did'
COMMENT  = 'Best is often a better view of the object.'
PLATE    =                266 / spectroscopic plate
```

图 5-8 FITS 头部中的标准型注释

b) 空格型：

常见的空格型注释的格式如图 5-9 和图 5-10 所示。左侧前 8 列仍使用“COMMENT”等关键字，不使用等号，通过空格分隔出右面的内容，右面并不需要使用单引号括起来。但是中间的空格数目并不确定，一些注释内容从第 9 列开始，一些从第 11 列开始。

```

EXTEND = T / FITS dataset may contain extensions
COMMENT FITS (Flexible Image Transport System) format defined in Astronomy and
COMMENT Astrophysics Supplement Series v44/p363, v44/p371, v73/p359, v73/p365.
COMMENT Contact the NASA Science Office of Standards and Technology for the
COMMENT FITS Definition document #100 and other FITS information.
    
```

图 5-9 FITS 头部中的空格型注释（一）

```

-----
DATE = '2001-05-22T15:54:31' / Date FITS file was generated
IRAF-TLM= '10:54:22 (22/05/2001)' / Time of last modification
COMMENT M67, 40 sec, filter B, 22:20:19UT, 29/12/2000
DATE-OBS= '2000-12-29'
    
```

图 5-10 FITS 头部中的空格型注释（二）

c) 空白型:

还有一类注释如图 5-11 所示，这一类在前 8 列并没有任何关键词出现，但是第 8 列之后会有一行或连续多行内容，一般也是对数据的补充说明。本文中称其为“空白型注释”。

```

LXTRASYM= 'NO' / Asymmetrical profile in extraction
LXTRCNTR= 50.9 / Center line of extracted spectrum
LFLUXAVE= 56.6 / Average flux (FNs)
COMMENT *
COMMENT * THE IUE VICAR HEADER
COMMENT *
COMMENT IUE-VICAR HEADER START
      895 895 768 768 1 1 011125637 +101 1 C
11057* 3*IUESOC * * * 660* * * * * * * * * * 2 C
LWP 25637, NGC 4151, 11 MINUTE EXPO, LO DISP, LG APER 3 C
ERRORS AT OFFSET STAR: EX,EY = (0,10) 4 C
      5 C
      6 C
OBSERVER: CRENSHAW, ID: AGORM, 31 MAY 1993, DAY 151 7 C
      8 C
      9 C
93151093538* 10 * 218 *OPS2PR16*093619 SCAN READLO SS 1 G3 47 * 10 C
063556 CAMINIT *093637 X 53 Y 71 G1 97 HT 106 * 11 C
063601 25636 18712 47775 1181 *093610 * 12 C
070527 S/C MANEUVERING *093636 * 13 C
    
```

图 5-11 FITS 头部中的空白型注释

d) 斜线型

除了每行的“关键词-注释”对后面的对该项属性的注释说明外，还有一些也以“/”开头的注释，但是它们并不针对某个具体属性进行说明，这些注释的后面通常有一组具有一定关联的关键词，因而这种注释一般是这一组关键词的说明，本文将这一类归为“斜线型注释”，如图 5-12 所示。


```

          / APERTURE POSITION
RA_APER1= 0.1826357301526E+03 / right ascension of the aperture (deg)
DECAPER1= 0.3940588837258E+02 / declination of the aperture (deg)

          / EXPOSURE INFORMATION
EQUINOX = 'J2000          ' / equinox of the celestial coordinate system
SUNANGLE= 0.1225114E+03 / angle between sun and V1 axis (deg)
MOONANGL= 0.1191039E+03 / angle between moon and V1 axis (deg)
SUN_ALT  = 0.4515910E+02 / altitude of the sun above Earth's limb (deg)
FGSLOCK = 'COARSE          ' / commanded FGS lock (FINE,COARSE,GYROS,UNKNOWN)

```

图 5-12 FITS 头部中的斜线型注释

鉴于这些注释内容往往也提供一些重要的信息，包括文件的修改历史、作者和生成文件的软件等，因此有必要将它们也抽取出来作为元数据的一部分。

这类注释信息的关键词包括“COMMENT”和“HISTORY”等，对于空白型和斜线型的注释，需要特别为其约定一个关键词，这里使用的是“MISC”。

而在保存这类注释信息的值时，可以注意到很多注释是跨行的，这是因为注释的信息量大，需要多行才能表述清楚。为完整保存这样的大文本，在遇到这种注释类型时，对同一个关键词（如“HISTORY”）按照它们在文中出现的先后次序，将其出现的各行信息连接起来，各行之间以回车换行符以及空格隔开，最终每个注释类型的关键词将各自拼接出一段完整的内容，并且保持了其在文中的换行、分段等格式，当需要还原成文中的信息时，这是很重要的。在 MySQL 数据库中，类型“text”就可以保存像回车换行这样的格式信息，对注释类型的关键词使用“text”类型就可以完成对注释内容和格式的完整保存。

◆ 附加信息

FITS 文件主头部中的元数据映射到数据库的字段后，还需要加上一些附加信息。因为数据库中同一张表内不应有重复记录，所以加上一列具有唯一性的主关键字是十分必要的。

此外，为了有效标识每条记录，可以将各记录所对应文件的相关信息也纳入元数据表中。根据目前天文台数据归档的情况，由于数据来源的广泛性和复杂性，文件名并不一定能唯一地标识一个数据文件，但是由于天文数据有比较严格的分类和归档要求，因此在整理之后的文件目录结构上可以很好地体现数据文件的来源、类型等信息，并且一定意义上已经具备了唯一性地标识文件的功能。因此可以将文件的路径信息作为元数据表中的附加信息字段。考虑到文件数据资源在物理存储的角度上看是可以移动、备份的，而数据文件之间的目录结构却是相对固定的，那么保存文件的相对路径将更具有现实意义。这样一方面具备了文件标识的功能，另一方面不会受限于物理存储的变动，有效屏蔽了一定的底层细节。

有必要定义这样一个“FITS 数据根目录”的变量，即所有这些 FITS 数据文

件的根目录。在需要确定某个 FITS 文件的完整路径时，将文件的相对路径和“FITS 数据根目录”的路径进行拼接即可。文件的相对路径、绝对路径以及 FITS 数据根目录这三者，只需要保存其中两个就可推出另一个。通过将“FITS 数据根目录”的值保存于配置文件或其他介质，就可以与当前元数据表的设计一起，具备从数据库记录定位到实际物理文件的功能。在 FitHAS 的当前实现方案中，在元数据表中存储的元数据和附加信息字段，只是一个最小的子集，在将来需要的时候，可以针对不同的应用场景进行扩充。

(2) 文件名模式过滤器

在进行 FITS 文件元数据的批量导入时，允许用户利用向导界面上的工具生成或手工输入一个文件名的模式过滤器，这使得在已选的文件或目录中只有文件名符合指定模式的那些文件才可以进行下一步的操作。

这里采用的表达式规范同 java 正则表达式一样，其中主要规则如表 5-3 所示。

表 5-3 文件名模式过滤器的特殊字符

字符	说明
.(点)	任意字符
\d	一个数字: [0-9]
\D	一个非数字: [^0-9]
\s	空白字符 [\t\n\x0B\f\r]
\S	一个非空白字符 [^\s]
\w	一个单词中可出现的字符 [a-zA-Z_0-9]
\W	一个非单词字符: [^\w]
X?	X, 一次或零次
X*	X, 零或多次
X+	X, 一或多次
X{n}	X, 刚好 n 次
X{n,}	X, 至少 n 次
X{n,m}	X, n 次到 m 次
XY	X 紧随着 Y
X Y	X 或 Y
(X)	X, 括号里的部分可组成一个组

5.4 FitHAS 系统实现

5.4.1 包结构

◆ 核心逻辑层

fitsfile 包: FITS 文件解析处理

dbsupport 包: 数据库操作支持

exception 包: 自定义的一些异常

util 包: 包含配置文件读写、压缩解压缩等通用方法

◆ 用户界面层

ui 包: 主要是 wizard 和 config 两个部分。

wizard 包: 包括 4 个工作向导, 分别是 FITS 头部查看向导、单一文件元数据入库向导、多文件元数据入库向导、数据库浏览向导。

config 包: 配置文件的读写

5.4.2 图形界面的实现

系统使用 java 的 swt/JFace 技术实现图形化界面, 包括图形化的向导、提升对话框、按钮、菜单等控件。使用的主要控件如下:

◆ 主页面

通过以下代码可以构成一个基本的图形化界面的框架

```
final Display display = Display.getDefault();
final Shell shell = new Shell();
shell.setSize(644, 463);
shell.setText("title1");
shell.layout();
shell.open();
//页面内容
while (!shell.isDisposed()) {
    if (!display.readAndDispatch())
        display.sleep();
}
```

◆ 向导

通过继承 `org.eclipse.jface.wizard.Wizard` 得到 `MyWizard` 类，可以控制何时前进、后退，以及向导中的每一步所执行的操作等。系统的 4 个向导模块均继承自 `MyWizard` 类，通过设置不同的向导页面和流程，完成各自的功能。

◆ 向导页

通过继承 `org.eclipse.jface.wizard.WizardPage` 得到 `MyWizardPage` 类，使每页都有自己的 id 号，从而有效地实现在向导中的翻页控制。向导中的每一步都是继承自 `MyWizardPage` 的一个子类（向导页）的实例，每个向导页可以有不同的页面布局、控制逻辑、页面输入框的有效性验证等。通过多个向导页的组合就可得到需要的向导。

◆ 按钮和菜单

按钮可以使用 `org.eclipse.swt.widgets.Button` 控件。菜单可以使用 `org.eclipse.swt.widgets.Menu` 和 `org.eclipse.swt.widgets.MenuItem` 控件。

5.4.3 异常的处理

为使系统具有一定的容错性，需要在执行文件元数据的批量入库时，能够确保程序不会因为遇到无法处理的文件或数据库操作失败等原因而被中断。这可以使用 java 的异常处理机制并结合一定的业务逻辑来实现。

◆ 异常原因

FITS 文件元数据入库的处理过程中出现的异常，主要有以下几种情况：

- ✧ 文件损坏、I/O 操作失败等
- ✧ 文件的主头部结构和所要导入的元数据表不一致，且是文件中的某个或某些关键词在元数据表中没出现
- ✧ 当前要导入的记录已在目标表中存在
- ✧ 文件没有位于“FITS 数据根目录”下
- ✧ 数据库连接异常及其他一些 `SQLException`

一般地，当需要导入的记录已存在于目标表时，将导致重复记录异常，这时，需要依据用户所选的处理重复记录的策略进行相应处理：跳过、更新、或者清空表再重新导入，如能正常完成处理过程则可以认为这次操作成功或者相当于说已经成功导入了这个记录。而其他大多数异常是因某个文件中的元数据不能被导入数据库而起或将导致入库的失败。这些异常会导致正常流程的中断，所以应在正确的地方对其进行捕获并做一定的处理，而不是简单地抛出异常和退出程序。

◆ 异常的处理

在对 FITS 文件头部元数据批量归档入库的过程中, 使用了递归的方法, 逐层向下查找子目录和文件并依次进行处理。在这种逐层调用的情况下, 当一个文件的处理过程中出现了异常, 在将控制权交回父一级目录后, 程序并没有退出, 而是继续对下一个文件进行处理。文件递归访问流程如图 5-13 所示。

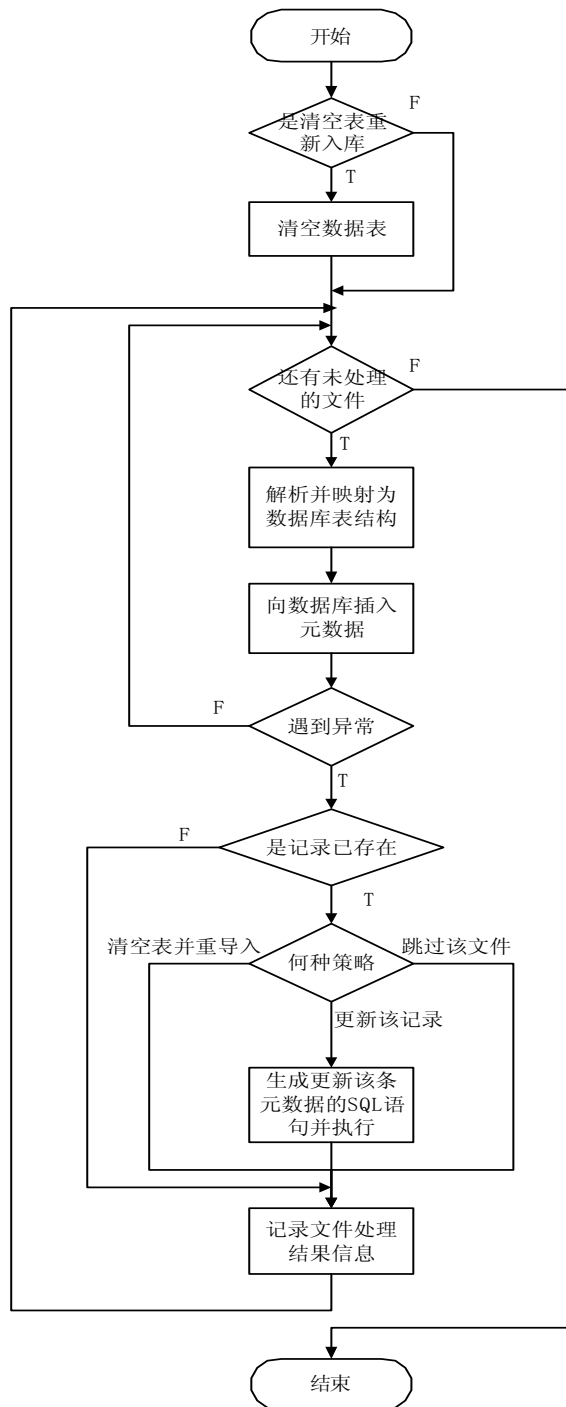


图 5-13 异常处理的流程

5.5 运行结果

系统支持 linux 和 windows 平台，需要 jre1.4 及以上版本的支持。
系统启动后，会看到如图 5-14 所示的界面。



图 5-14 FitHAS 主界面

在浏览 FITS 头部向导中，FITS 头部信息显示如图 5-15 所示。

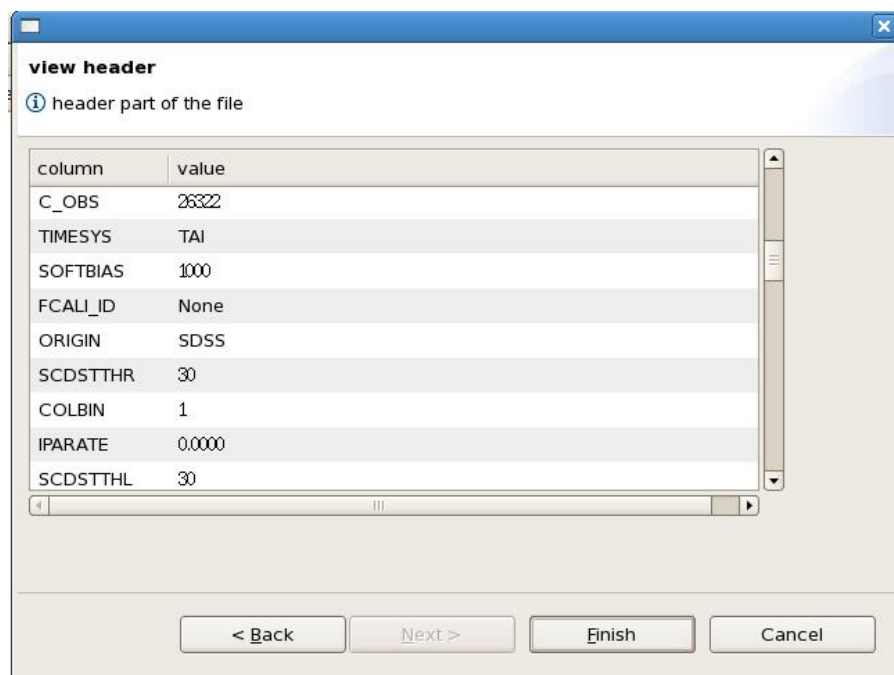


图 5-15 浏览 FITS 文件头部信息的界面

在对多个文件进行批量归档入库时,可以在向导界面上选择多个文件或者一个文件目录,如图 5-16 所示。选中的文件目录下的所有子文件夹和文件将被递归地导入元数据库。

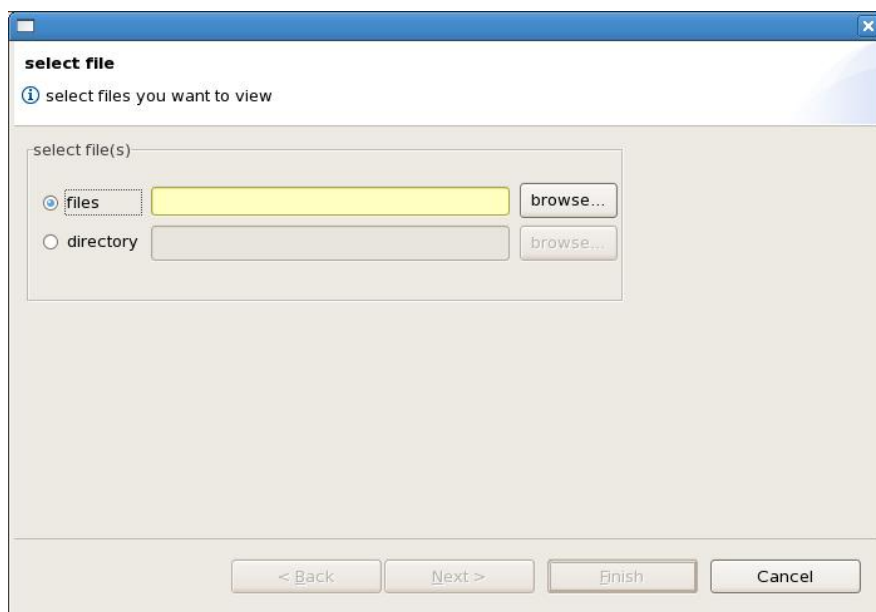


图 5-16 选择文件的界面

如图 5-17 所示,在执行批量文件归档的向导中,允许用户设置文件名的模式匹配过滤器,用户可以直接填写模式的表达式,也可根据下方的提示生成表达式,并且可以选择文件类型是普通 FITS 文件还是压缩的 FITS 文件,则上一步选择的范围中只有匹配当前模式的文件才能执行元数据入库的操作。

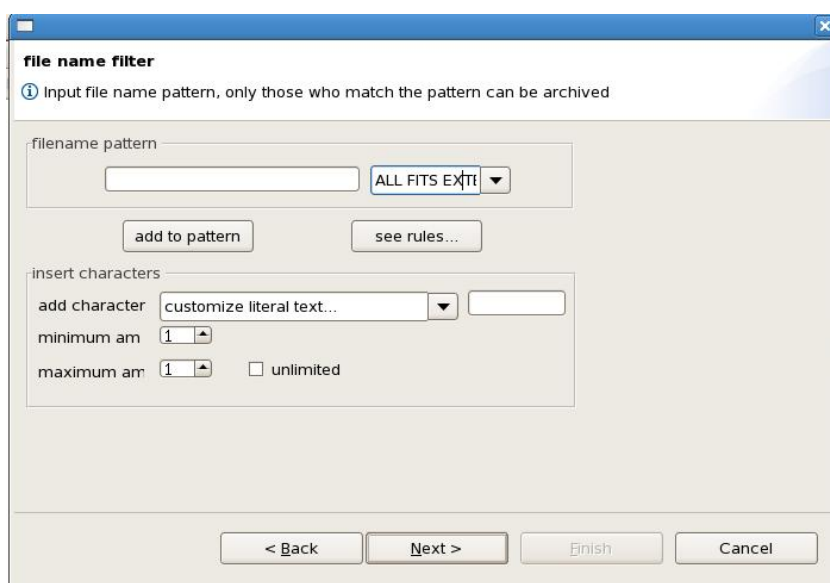


图 5-17 设置文件名过滤规则的界面

在数据库设置页面，用户可以从下拉列表中选择一个已有的元数据表，或者也可以根据已选择的文件新建一个元数据表，如图 5-18 所示。

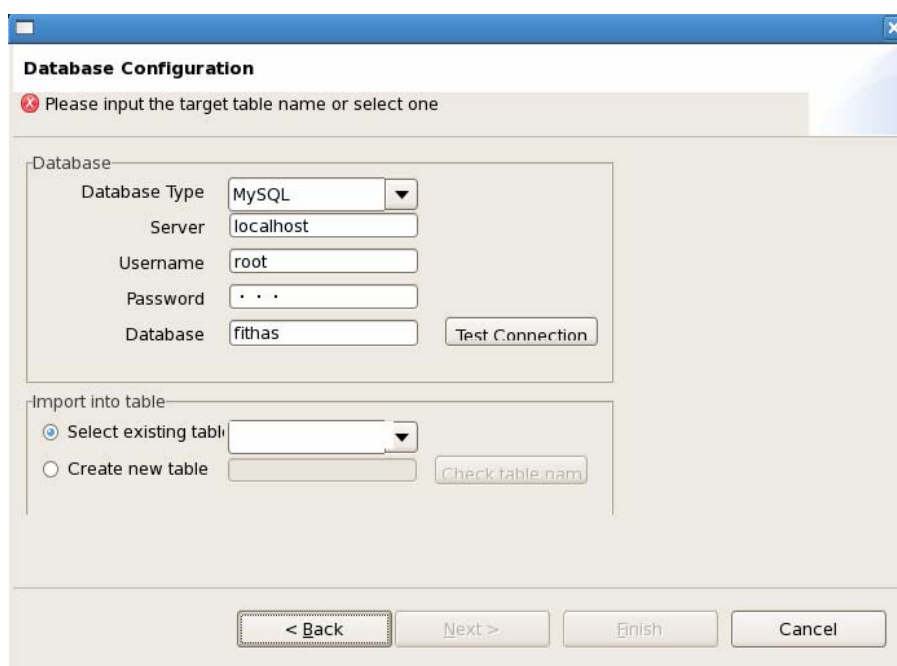


图 5-18 数据库设置的界面

在执行归档操作之前，需要设置对重复的元数据记录的处理策略，如图 5-19 所示。

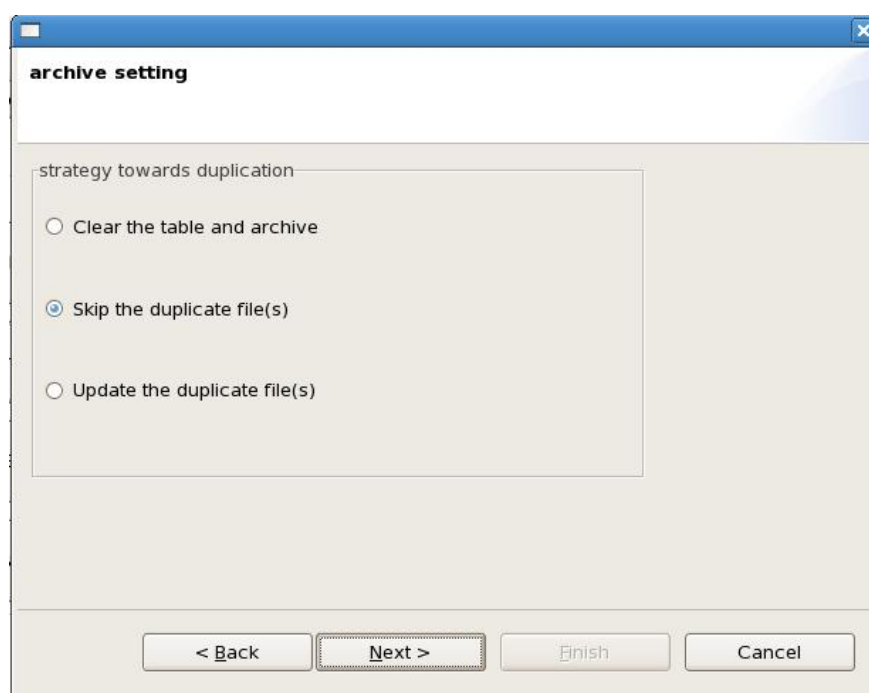


图 5-19 避免元数据记录重复的策略设置界面

元数据归档操作完成后，系统将显示操作结果报告信息，包括成功/失败的文件数目、文件路径，以及失败原因等，如图 5-20 所示。

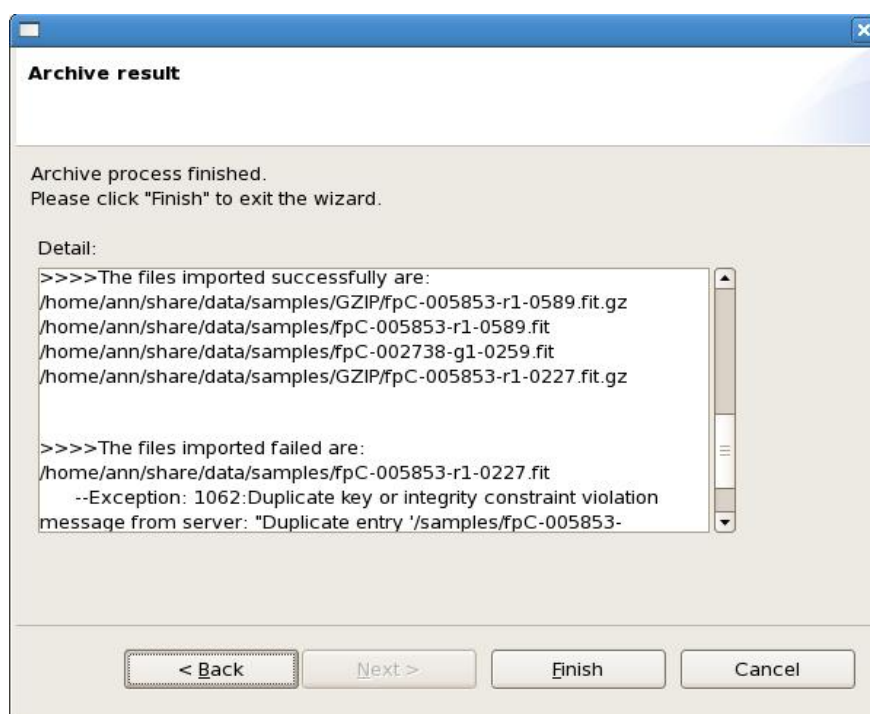


图 5-20 入库结果界面

用户可以通过元数据表浏览向导查看归档结果，如图 5-21 所示。

record_id	path	SCI_SN	XCHI2	DUSTD	TAIHMS	DUSTC	FBADPIX
1	/SDSS/spec	0.00000000	1.26351889141	451.551767621	04:10:12.13	18885.672246	0.0105899
2	/SDSS/spec	0.00000000	1.26351889141	451.551767621	04:10:12.13	18885.672246	0.0105899
3	/SDSS/spec	0.00000000	1.26351889141	451.551767621	04:10:12.13	18885.672246	0.0105899
4	/SDSS/spec	0.00000000	1.26351889141	451.551767621	04:10:12.13	18885.672246	0.0105899
5	/SDSS/spec	0.00000000	1.26351889141	451.551767621	04:10:12.13	18885.672246	0.0105899
6	/SDSS/spec	0.00000000	1.26351889141	451.551767621	04:10:12.13	18885.672246	0.0105899
7	/SDSS/spec	0.00000000	1.26351889141	451.551767621	04:10:12.13	18885.672246	0.0105899
8	/SDSS/spec	0.00000000	1.26351889141	451.551767621	04:10:12.13	18885.672246	0.0105899
9	/SDSS/spec	0.00000000	1.26351889141	451.551767621	04:10:12.13	18885.672246	0.0105899
10	/SDSS/spec	0.00000000	1.26351889141	451.551767621	04:10:12.13	18885.672246	0.0105899
11	/SDSS/spec	0.00000000	1.26351889141	451.551767621	04:10:12.13	18885.672246	0.0105899
12	/SDSS/spec	0.00000000	1.26351889141	451.551767621	04:10:12.13	18885.672246	0.0105899
13	/SDSS/spec	0.00000000	1.26351889141	451.551767621	04:10:12.13	18885.672246	0.0105899

图 5-21 数据库浏览界面

第六章 总结与展望

6.1 工作总结

本文对比分析了当前的几种主要存储技术,并根据中国虚拟天文台对海量数据存储的要求,提出一种切实可行的存储方案。

通过对天文数据访问方式的讨论,本文采用通过数据库保存天文数据文件的元数据,有效提高了对海量数据文件的访问速度,并通过实例论证该方案的可行性和有效性。在此基础上进一步提出了一种虚拟天文台环境下的天文数据的分层模型。

本文还在海量天文数据的存储和高效访问研究的理论基础上,设计并实现了一套 FITS 文件元数据归档管理系统 FitHAS,以解决 China-VO 所面临的从海量数据中构造元数据层的难题,在实际使用过程中收到了良好的效果。

6.2 展望

目前的 FitHAS 系统只能访问本地文件,在下一步的工作中,可以集成 ftp 功能,使程序能够访问异地文件。并且进一步增加对直接读取异地文件的支持,从而可以只将解析得到的 FITS 文件头部的信息在网络上传输以有效减小对网络带宽的影响。

进一步完善 FITS 文件元数据到数据库的映射规则,建立统一的标准,以使现有的应用和将要开发的服务都可以很好地支持储存元数据的表,从而将各软件更容易地集成到虚拟天文台中去。

实现 FitHAS 的命令行方式的版本,从而可以更有效地将系统集成到虚拟天文台的服务之中。通过编写批处理脚本等方式,实现海量数据管理的自动化日常维护。

参考文献

- [1] 黄力, 杨志良, 现代天文教育和研究平台——虚拟天文台, 北京师范大学学报(自然科学版), 2005, 41(03): 300~302
- [2] 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, 计算机应用研究, 2006, (02): 125~128
- [3] 崔辰州, 赵永恒, 中国虚拟天文台体系结构, 天文研究与技术, 2004, 1(2): 140~151
- [4] China-VO, www.china-vo.org
- [5] LAMOST, www.lamost.org
- [6] 崔辰州, 中国虚拟天文台系统设计: [博士学位论文], 北京: 国家天文台, 2003
- [7] 王达, 网管员必读——服务器与数据存储, 北京: 电子工业出版社, 2005, 305~537
- [8] 时成阁, 网络存储导论, <http://www.dostor.com/f/NSDL/>, 2006
- [9] Robert Spalding, Storage Networks: The Complete Reference, McGraw-Hill Osborne Media, 2003, 928
- [10] 赵文辉, 徐俊, 周加林等, 网络存储技术, 北京: 清华大学出版社, 2005, 14~23
- [11] Thomas M. Ruwart, Disk Subsystem Performance Evaluation: From Disk Drives to Storage Area Networks, 17th IEEE Symposium on Mass Storage Systems / 8th NASA Goddard Conference on Mass Storage Systems and Technologies, 2000
- [12] John Vacca, 存储区域网络精髓(郭迅华等译), 北京: 电子工业出版社, 2003, 324~343
- [13] 颜维龙, 盖杰, 武港山等, 面向网络的全文检索中索引文件的组织, 计算机应用研究, 2002, (11)
- [14] 宋雯斐, 全文检索研究与系统实现: [硕士学位论文], 北京: 北京师范大学, 2004
- [15] 曹右琦, 孙茂松, 一种面向文件的高效检索方法, 中国中文信息学会二十五周年学术会议, 北京, 2006
- [16] 崔辰州, 虚拟天文台——网格技术最好的试验场, 中国计算机报, 2004
- [17] 崔辰州, 赵永恒, 赵刚等, 虚拟天文台的技术进展, 天文学进展, 2002, 20(4): 302~311
- [18] 崔辰州, 天文学的新革命——虚拟天文台, 天文爱好者, 2001, (05)

- [19]张彦霞, 赵永恒, 虚拟天文台: 科学、工具及应用, 天文学进展, 2006, 24 (03): 189~199
- [20]刘鹏, 网格概念的界定, <http://www.chinagrid.net/grid/paperppt/GridConcept.pdf>, 2003
- [21]Ian Foster, Carl Kesselman, The Grid: Blueprint for a New Computing Infrastructure, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999
- [22]赵永恒, 中国虚拟天文台——任务、特点、方案, 2002
- [23]LAMOST 工程指挥部科学部, LAMOST 科学目标 (2006 年版), 2006
- [24]SDSS, <http://www.sdss.org/>
- [25]赵永恒, China-VO 的发展, www.china-vo.org/cn/events/0309BJ/China-VO_dev.ppt, 2003
- [26]Microsoft, Microsoft Exchange Server 2003 High Availability Guide, <http://www.microsoft.com/downloads>, 2006
- [27]Kaladhar Voruganti; Prasenjit Sarkar, An Analysis of Three Gigabit Networking Protocols for Storage Area Networks, Performance, Computing, and Communications, 2001. IEEE International Conference on., Phoenix, AZ, USA2001, 259~265
- [28]中国科学院计算机网络信息中心, 科学数据库超级数据服务器运行情况报告, <http://www.sdg.ac.cn/document/yunxing.zip>, 2005
- [29]刘燕之, 瞬间搜索海量数据, 每周电脑报, 2004, 24~25
- [30] (白皮书) 迈向国家虚拟天文台: 科学目标、技术挑战和实施计划 (草案), http://www.china-vo.org/docs/nvo_white_cn.PDF, 2000
- [31]VOTable, <http://vizier.u-strasbg.fr/doc/VOTable/>
- [32]桑健, 赵永恒, 崔辰州, 中国虚拟天文台数据访问服务, 天文研究与技术, 2004, 1 (3): 216~228
- [33]Geoff Brumfie, Astronomy: The Heavens at Your Fingertips, Nature, 2002, 420(6913): 262~264
- [34]柯大荣, 赵永恒, FITS 基本格式及其扩展, <http://www.lamost.org/~yzhao/doc/fits.html>, 1993
- [35]柯大荣, 赵永恒, 一种图象传输系统及其 FITS 数据基本格式, 现代图书情报技术, 1994 (2)
- [36]Francine Berman, Geoffrey C.Fox, Tony Hey, 网格计算: 支持全球化资源共享与协作的关键技术 (都志辉等译), 武汉: 华中科技大学出版社, 2005
- [37]汪燕芳, 赵可新, 卢晓猛, 用 VC 实现 FITS 和 BMP 格式的转换, 微计算机信息, 2006, 22 (4)
- [38]汪燕芳, 基于 FITS 格式的天文图像的识别与跟踪: [硕士学位论文], 北京: 北京科技大学, 2005

- [39]R. J. Hanisch, A.Farris, E.W.Greisen, et al, Definition of the Flexible Image Transport System (FITS), *Astronomy & Astrophysics*, 2001
- [40]IAU FITS Working Group, Definition of the Flexible Image Transport System (FITS), 2005
- [41]D.C. Wells, E.W.Greisen, R.H.Harten, Fits: A Flexible Image Transport System, *Astronomy & Astrophysics supplement series*, 1981, 44: 363~370

发表论文和科研情况说明

发表的论文:

- [1] 王刚, 孙济洲, 李文, “一种启发式双代号网络图自动生成算法”, 《计算机应用》, 2007 年 3 月

参与的科研项目:

- [1] 支持 ERP 开发的中间件与组件系统 (编号: 04310971R), 天津科技攻关专项
- [2] 天津市 12319 城建服务热线系统
- [3] BATC 图像数据发布系统, 中科院国家天文台委托项目

致 谢

本论文的工作是在我的导师孙济洲教授的悉心指导下完成的，孙老师严谨的治学态度和科学的工作方法给了我极大的帮助和影响。在此衷心感谢几年来孙老师对我的关心和指导。

科研工作的发展和论文的顺利完成，是与于策老师的指导分不开的，于策老师对我的论文的撰写提出了很多宝贵的意见，在此向于策老师表示衷心的感谢。

感谢国家天文台的崔辰州博士，他对我的科研工作给予了很多指导，并且在工作和生活上给予了我很大的关心和照顾，使得科研工作可以顺利的开展。

在实验室工作及撰写论文期间，张坤龙老师对我的论文提出了宝贵的意见，在此向张老师表达我的感激之情。

科研过程中还得到了实验室的乌晓峰、周宁、王刚、武华北几位师兄和何团伟、徐祯、亓大志以及天文台的刘超、罗宇等同学的帮助，同时我宿舍的同学在此期间也给予了我最大的理解和支持，在此一并表示感谢。

另外也感谢我的家人，他们的理解和支持使我能够在学校专心完成我的学业。