

国家卫星气象中心气象卫星资料存档系统*

钱建梅 郑旭东

(国家卫星气象中心,北京 100081)

摘 要

论述了国家卫星气象中心的气象卫星资料存档系统构成、存储数据的编目组织、应用程序的逻辑关联结构、层次框架结构及组织结构,介绍了应用程序主要功能和流程图,对存档系统建设中的经验进行了总结。

关键词:气象卫星 编目 存储

引 言

数据是信息时代的最重要的资源之一。实现数据的格式标准化、处理规范化、保存永久化、访问自动化是气象卫星资料处理的重要技术之一,也是21世纪气象卫星资料应用领域的重大课题之一。

国家卫星气象中心存档系统1998年开始建设,1999年投入正式业务运行以来,气象卫星资料的存档采用自动磁带库作为主要的资料存储设备,改变了过去卫星资料人工存档,检索提取脱机进行的状况,实现了联机实时、自动化资料存档、检索与提取,使卫星资料的存档和使用向现代化大大推进了一步。

1 气象卫星资料的特性与存储内容

累积性:气象卫星资料之庞大是由于每次接收的数据量大而且又是不断地日积月累。每天10至20GB的观测资料,一年累积的数据量为4至7TB。

永久性:气象卫星每天对地球大气与地表环境进行监测。通过对历史资料的反演、与当前资料的对比,可以发现大气与地表环境的改变,预测未来气候与环境的变化,因而是非常宝贵的资料,需要永久保存。

复用性:天气气候研究与应用服务部门经常需使用历史资料,如何使用户方便快捷地获取资料是一个非常重要的方面。

目前国家卫星气象中心资料存档的主要内容与存储量如表1所示。

2 存档系统构成

国家卫星气象中心现有存储系统由HPK570服务器、磁盘阵列、STK9360自动磁带库、VERITAS存储管理软件、SYBASE数据库和存储编目检索应用软件构成,存储系统

* 2002-09-17收到,2002-11-27收到修改稿。

表 1 目前国家卫星气象中心存档资料的内容与每天的存储量

卫星	内容	存档量
FY-1C	CHRPT 1B 数据集	$150 \text{ MB} \times 7 = 1.05 \text{ GB}$
	GDPT 1A.5 数据集	$55 \text{ MB} \times 16 = 1 \text{ GB}$
	产品	约 0.6 GB
FY-1D	CHRPT 1B 数据集	$150 \text{ MB} \times 7 = 1.05 \text{ GB}$
	GDPT 1A.5 数据集	$55 \text{ MB} \times 16 = 1 \text{ GB}$
	产品	约 0.6 GB
NOAA 17	HRPT 1B 数据集	$100 \text{ MB} \times 7 = 0.7 \text{ GB}$
	产品	约 0.1 GB
NOAA 16	HRPT 1B 数据集	$100 \text{ MB} \times 7 = 0.7 \text{ GB}$
	产品	约 0.1 GB
FY-2B	圆盘图存档资料	$100 \text{ MB} \times 28 = 2.8 \text{ GB}$
	产品	约 0.5 GB
GMS5	圆盘图存档资料	$100 \text{ MB} \times 28 = 2.8 \text{ GB}$
	产品	约 0.3 GB
EOS/ MODIS	0 级数据(PDS)	$800 \text{ MB} \times 4 = 3.2 \text{ GB}$
	1 级数据(HDF)	$1.2 \text{ GB} \times 4 = 4.8 \text{ GB}$
合计		约 23.7 GB

注: CHRPT: FY-1 卫星高分辨率图像传输(China High Resolution Picture Transmission);

GDPT: 全球延时图像传输(Global Delay Picture Transmission);

PDS: 产品数据集(Product Data Set);

HDF: 分层数据格式(Hierarchical Data Format)。

服务器、数据库、带库联接示意图见图1。其中两台 HPK570 服务器作为 STK 磁带库与 SYBASE 数据库的主控服务器互为备份,采用 HP 高可用软件,当一台服务器或某一类应用软件故障时,整机业务或故障业务自动切换到另一台服务器。STK 磁带库与 SYBASE 数据库分别采用浮动主机名与浮动 IP 地址,使服务器整机业务或某应用的切

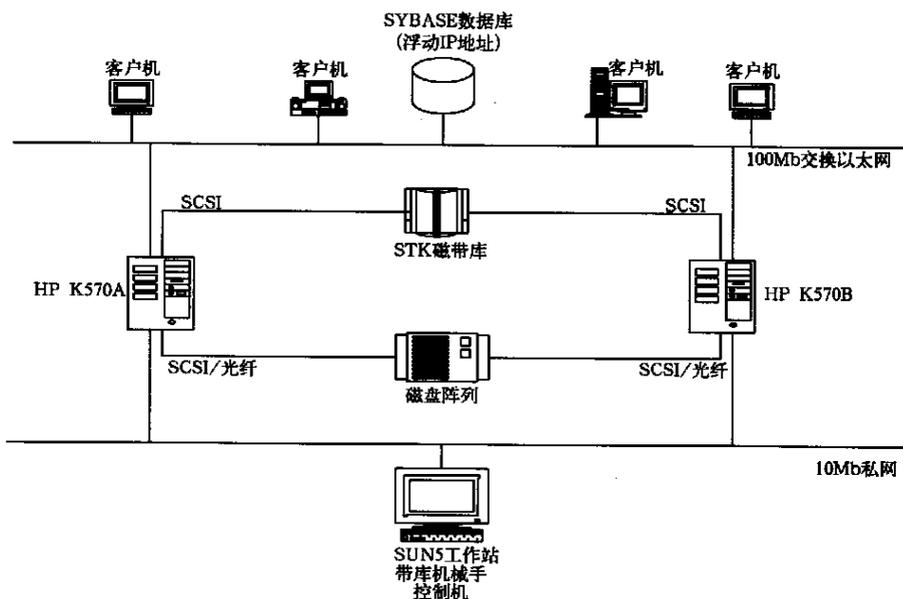


图 1 国家卫星气象中心存档系统服务器、数据库、带库联接示意图

换对磁带库与数据库的使用而言完全透明。

2.1 存储数据的编目与组织

从存储方式看,气象卫星资料的存储分为三级,在线数据、近线数据、离线数据。在线数据指存储在磁盘阵列中的数据,用户访问延迟时间最短,但存储数据量与期限受磁盘容量的限制。近线数据指存储在自动带库中的数据,存储数据量与磁带库的容量相关,保存期限视情况可定义为数月、数年或永久保存,用户访问该类数据时有从磁带读数据的时间延迟。离线数据指已出库上架的磁带,一般为长期或永久保存的数据,访问该类数据时,可以实现联机检索,但读数据时磁带需要重新入库,延迟时间较长。合理的三级数据存储组织,可以有效解决速度与容量的矛盾。

从数据编目角度看,可将气象卫星资料划分为结构化数据和非结构化数据。结构化数据指经提取的、能够对某具体的数据对象或过程控制进行抽象描述的关系型数据,它包括元数据、业务监控信息、配置管理参数等。非结构化数据泛指气象卫星资料和产品数据集,由文件系统、二级存储数据、离线数据构成。数据编目采用多级编目结构,结构化数据作为主编目,下设三级子编目,分别对应非结构化数据的文件系统目录、二级存储数据目

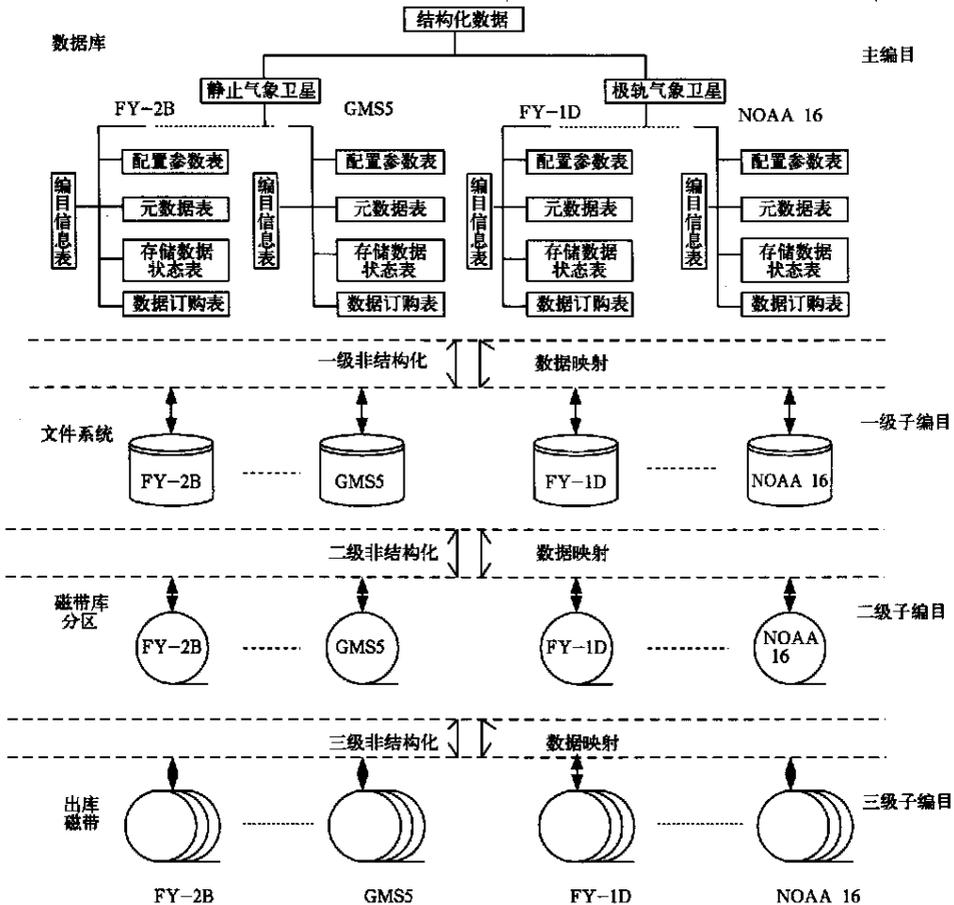


图 2 气象卫星资料的编目结构示意图

录和离线数据管理目录。使用数据库构建结构化数据主编目,文件系统目录构建非结构化数据一级子编目,自动磁带库的磁带分区目录构建非结构化数据二级子编目,已出库磁带数据目录构建非结构化数据三级子编目。客户通过一级非结构化数据映射,存储访问文件系统上的在线数据,通过二级非结构化数据映射,存储访问磁带库中的近线数据,通过三级非结构化数据映射,访问已出库的离线数据。主编目和一级、二级、三级子编目的编目信息均存放于数据库中。气象卫星资料的编目结构如图 2 所示。

2.2 应用软件结构

2.2.1 逻辑关联结构

存档系统的存储编目检索应用软件部分是基于 VERITAS 存储管理软件和 SYBASE 数据库系统由国家卫星气象中心的技术人员自行设计开发。应用软件分为两个部分:主机端部分和客户端部分。应用软件、SYBASE 数据库、VERITAS 存储软件以及磁带库之间的逻辑关联如图 3 所示。

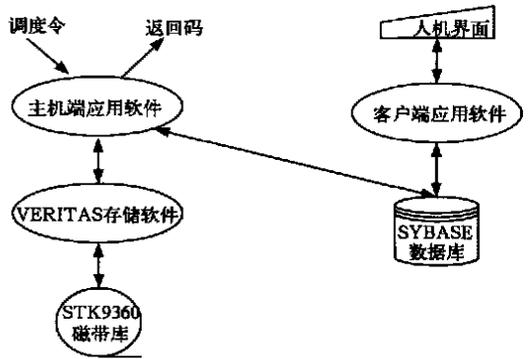


图 3 应用软件、SYBASE 数据库、VERITAS 存储软件以及 STK 磁带库逻辑关联示意图

2.2.2 层次框架结构

存储编目检索软件的层次框架结构是基于由存储检索服务器、数据库服务器和资源管理服务器以及一个或多个客户程序构成三层的三角形架构的客户/服务器模型,如图 4 所示。自动存储业务和用户的交互访问通过存储检索服务器进行。由于气象卫星数据量的巨大,数据分为结构化数据和非结构化数据两个层面进行编目管理。访问

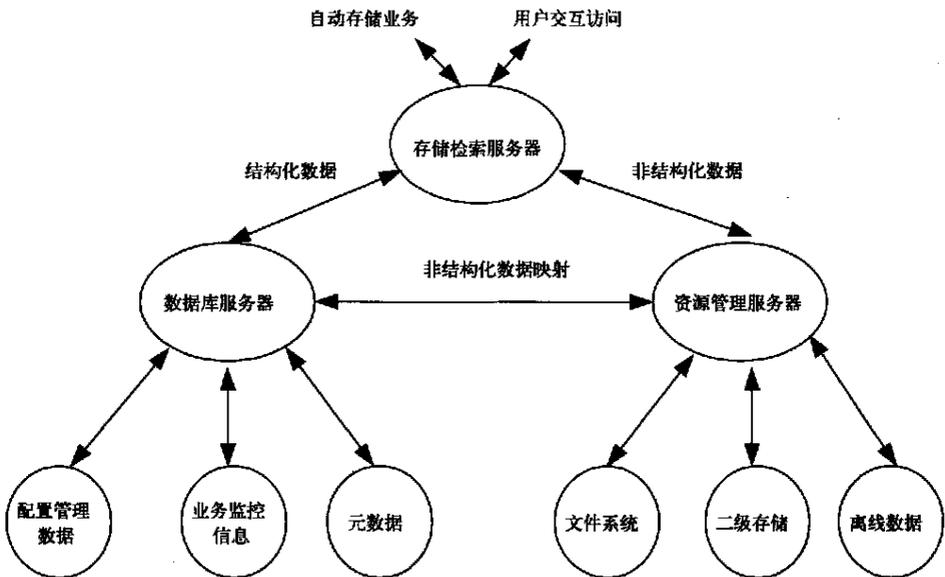


图 4 存储编目检索应用软件的层次框架结构示意图

非结构化数据时通过结构化数据进行逻辑化物理映射、监控和管理。

2.2.3 软件组织结构

存储编目检索软件由主机软件子系统和客户机软件子系统两部分构成,主机软件子系统主要完成资料的存储、编目、管理和下载,客户机软件子系统为用户提供各种已存储资料信息的检索、统计和订购处理。每个子系统含若干软件及模块,其组织结构如图 5 所示。

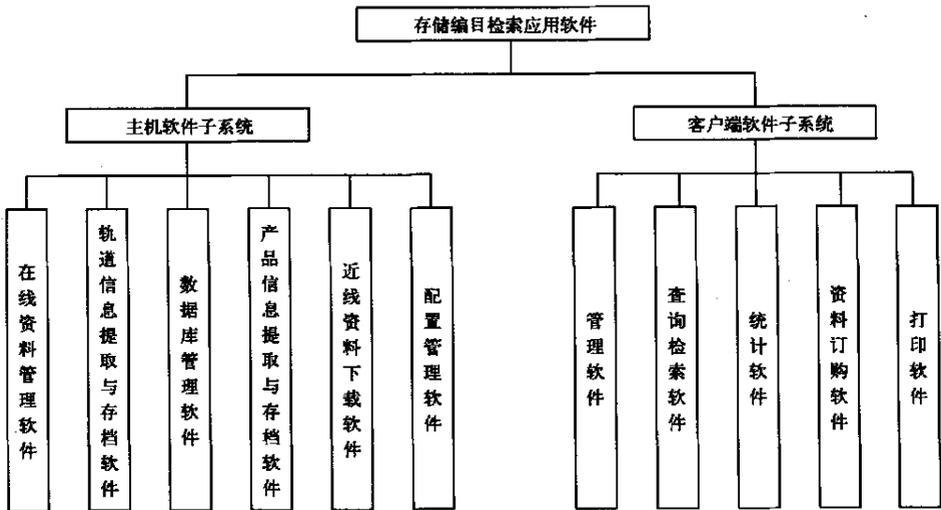


图 5 存储编目检索软件组织结构

2.3 应用软件主要功能

主机软件主要功能:(1) 调度令解析;(2) 元数据提取、入库;(3) 存档资料、产品的质量检查、存档;(4) 存档作业状态信息提取、入库;(5) 卫星、存档数据、磁带库资源的配置管理;(6) 数据库库表建立、关键字定义、数据库常规操作功能实现;(7) 客户端数据请求的解析、资料产品的获取;(8) 在线资料管理。

客户端软件主要功能:(1) 为不同用户提供所授权的信息;(2) 用户管理;(3) 按各种条件查询检索资料内容;(4) 资料订购与管理;(5) 配置项管理;(6) 统计、打印及其它常规功能。

2.4 系统特点

(1) 配置参数管理 通过设立配置参数的方法对卫星参数、资料产品类别、磁带库资源等进行配置,存储时根据配置参数自动实现存档。当增加新的卫星、产品或业务变更时,只需要增加或修改相应的配置参数即可,而无需修改程序。使用配置参数的方法也大大减少了编程量。

(2) 使用 SYBASE 的外部调用功能完成客户端到服务器的数据请求 服务器与客户端之间的通信可以有很多方法,SOCKET、MQ 消息等等,但要做到稳定运行需要非常完善的故障处理机制。使用 SYBASE 外部调用功能实现数据请求,大大减少了程序的复杂性,并且不受网段设置等条件的限制。

(3) 应用软件与商用软件的无缝连接 一般商用软件都是自成体系,它给用户提供的信息是有侧重的、有授权限制的、与用户业务系统分离的,操作员在进行业务监视、质量检查时很不方便也不够用。我们从 VERITAS 存储软件的并发作业信息流中获取存档和资料下载的作业状态,与资料的元数据一起提供给操作人员进行监视和质量检查,使得业务系统成为一个有机的整体。

(4) 逻辑化物理映射 使用 VERITAS 存储软件作近线存储时,它不返回所存数据的物理位置,资料存入磁带库就好比进了“黑匣子”,这给资料的深层次管理带来困难。比如由于磁带驱动器硬件故障造成磁带损坏需要对数据进行修复,某些特定资料需要进行复制等等。特别是当脱离 VERITAS 存储软件时,用户面对的是一堆磁带,而不是数据,不利于灾难恢复。我们从 VERITAS 存储软件的专用数据库中取出数据存储的物理位置,并作为结构化数据保存在数据库中,实现了存储数据逻辑化的物理映射。在二级存储设备更新时的大批量气象卫星资料转储中,这一逻辑化的物理映射信息发挥了重要作用。

(5) 后延性 存档系统的后延性指两个方面:存储量大量增长后的系统开销和存储数据与运行系统的相关性。有的存储软件在存储量比较小时,具有非常灵活方便的管理功能,但当存储量增大到一定程度时系统开销非线性增长,出现存取瓶颈。存储数据与运行系统的相关性是指所存储的信息以当前运行系统某些特定工具作为数据载体,以至当系统更新淘汰时很难对这些数据进行恢复和转储。特别是永久保存的数据,应尽量避免此类问题的发生。

3 结束语

国家卫星气象中心的存档系统将数据存储、数据可用性、集中式资源管理和主动策略管理技术结合在一起,保证了数据的急剧增长不影响软件有效存储、访问和管理数据的能力。该系统自启用以来,已存储 FY-1 系列、FY-2 系列、NOAA 系列、GMS 系列、EOS/MODIS 系列气象卫星各种资料产品 10 TB 多,提取结构化数据 50 多万个条目,为各类用户提供检索和资料下载服务 5 万多次,在天气预报和研究、灾情和环境监测中发挥了显著作用。

THE SATELLITE DATA ARCHIVE SYSTEM OF NATIONAL SATELLITE METEOROLOGICAL CENTER

Qian Jianmei Zheng Xudong

(*National Satellite Meteorological Center, Beijing 100081*)

Abstract

Data is one of the most important resources in the information times. Because of the immensity capacity of meteorological satellite data, it is significant that how to save it permanent, access it automatically and process it standardization, and it is also one of the important tasks in the meteorological satellite data application field in the 21st century. More than 10 TB data has been stored since the system running including raw data and various products of FY-1 series, FY-2 series, NOAA series, GMS series, EOS/ MODIS series. The structure of satellite data archive system, the catalogue of storage data, the logical relationship of the hierarchical structure and the organizing of application software are discussed. The primary function and the flow chart of application software are also introduced.

Key words: Meteorological satellite Data catalogue Information storage