

# 气候资料分布式数据库总体框架探讨\*

花灿华

(国家气象中心, 北京 100081)

## 提 要

文章着重阐述了系统目标, 要求在数据表示、数据结构、应用接口和用户界面等方面实行统一规范, 以保证数据的一致性和互换性, 从而使气象系统及非气象系统的相关系统均可纳入一个总的框架, 相互连接, 彼此调用。同时, 阐述了系统的结构。该系统在地理上是国家-区域-省-地气候资料数据库的多数据结构, 但在分布式数据库管理系统控制下, 在逻辑上是一个整体数据库。

**关键词:** 气候资料 分布式数据库 系统目标和结构

## 前 言

气候资料是经济建设、国防建设和科学研究不可缺少的重要依据。在工业布局、农业气候区划、气候资源开发, 乃至尖端科技试验等方面, 均有着广泛的应用。在气象科研和日常业务中, 则更是使用频繁, 不可或缺。因此, 加强气候资料的管理, 提高使用效率是气象工作的一个重要课题。

随着近代科学技术的发展, 气候研究内容不断扩大和深化, 气候资料的概念和内涵已远远超出人们通常所认为的常规观测资料的范围, 而应包括整个气候系统, 即大气圈、海洋圈、冰雪圈、陆地和生物圈的有关资料。人们已认识到, 对气候系统的监测和预测, 必须与全球气候观测系统(GCOS)的气候资料管理系统相结合, 并将逐步发展成为地球观测系统(EOS)世界网中的重要组成部分。面对客观变化的现实, 研究采用先进的客户/服务器体系结构和分布式数据库技术, 是适应时代发展的需要。

考虑到气候资料的种类繁多, 各种资料的数据量庞大, 而且随着用户使用特点的不同, 使要素项目、序列长度、时空范围迅速扩展。因此, 反映气候系统运动状态的气候资料数据库必然是多数据库联合体, 而且也必然是资料库(文件管理)和数据库相结合的体系结构。本文将着重探讨气候系统基本信息的宏观分布, 建立一个统一的规范和标准, 保证数据的一致性, 以适应广泛的应用, 并准备通过与气象系统以外的计算机和数据库连接的接口, 达到可与其它子系统相互对接, 彼此调用, 数据互换, 信息共享。

\* 1996-04-22-收到, 1996-08-12 收到修改稿。

## 1 气候资料分布式数据库的总体目标和建库原则

(1) 气候资料分布式数据库根据世界天气监视网(WWW)的分布式数据库(DDBS)概念和世界气候资料计划(WCDP)的气候资料管理(CDM)的基本原则,利用近代分布式数据库技术,对全球气候资料进行统一组织、管理和操作,可以网络互联、灵活扩展、高效检索,实现全国气象系统的资料共享,更好地为国家各部门和气象业务、科研提供资料服务。

(2) 气候系统的信息是多学科跨部门的,各部门的信息源不可能集中在一个总数据库内,各部门相对独立的应用系统仍需要各自的数据库来管理各自的信息。因此,全球气候资料根据气候系统多时间尺度运动和多子系统相互交叉的特点,按不同部门的资料分类建库。所以,气候资料分布式数据库必须是一个具有良好开放性的系统,易于实现与其它部门分布式数据库的连接,形成跨部门的数据库系统。

(3) 气候资料分布式数据库要求在数据表示、数据格式、数据结构、应用接口和用户界面等方面实行统一规范,遵从SQL(结构查询语言)和RDASP(远程数据访问服务和协议)OSI(国际标准),以保证数据的一致性和互换性,从而使气象系统及非气象系统的相关子系统均可纳入一个总的框架,相互连接,彼此调用。

(4) 根据气候资料服务和管理的需要,气候资料分布式数据库在全国气象系统范围内,分国家-区域-省-地四级管理,分布控制。建库的分布策略,按各级管理职能范围和产品,按统一规划分别建库。在尽量避免重复的原则下,允许有一定的冗余度,以求最大限度的检索效率和管理效益。同时按资料的共享范围和业务需求,对公用资料统一建库,集中管理;对专用资料分别建库,各自管理。

## 2 气候资料分布式数据库的基本结构

(1) 数据库系统在地理上呈现为国家气候资料库(NCDB)、区域气候数据库(RCDB)和省气候资料数据库(PCDB)三级自治的多数据库结构。但在分布式数据库管理系统的控制下,在逻辑上是一个整体数据库<sup>[1]</sup>。

(2) 各级系统均有形式类似、规模不同与实时库相匹配的历史库,气候资料历史数据库以及磁带库(数据集)综合构成。

(3) 与实时库相匹配的历史库以临时数据库方式与实时库结构表示,由磁带或其它载体作为其常驻方式。气候资料历史数据库主要按气候序列,即以测站为单位。磁带库(含入数据库和不入数据库的数据集)与数据库之间的公共目录管理,对入库数据而言,磁带库与数据库之间有转换接口。

(4) 实时库与气候资料历史数据库的集成结构如图 1 所示。

实时库(RDB)截断时间一般为 10~15 天。期内的实时检索在 RDB 上实现。超过截断时间的实时检索,其存取路径需由 RFA 从 TLMS 的数据集中进行重构。在重构的历

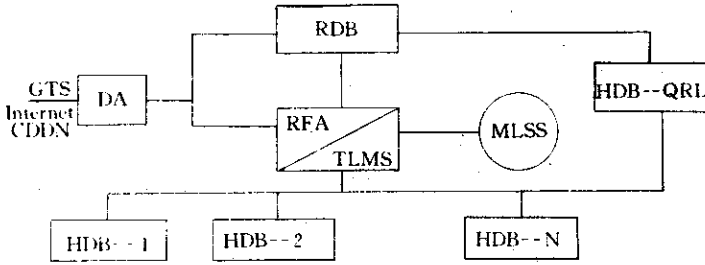


图 1 实时库与气候资料历史数据库的集成结构

Fig. 1 The integrated structure for real-time database and historical database of climate data

史库上进行检索(支持软件、功能与实时库相同). 对气候资料可从 RFA 中提取指定要素进入气候资料历史数据库(HDB), 执行数据库检索. 另外, 还建立一个准实时历史资料数据库(HBD-QRL), 这个库主要装有尚未加到气候历史资料数据库中的经某些加工后的实时库的数据. 这主要是为了能使这些数据及时地与气候历史资料数据库中的数据一起提供服务. 同时, 也是为了使这些数据及时地追加到气候历史资料数据库中. 因此, 这个库是作为正式入库前的历史资料数据的准实时历史资料数据库.

(5) 数据库与磁带库的集成结构如图 2 所示.

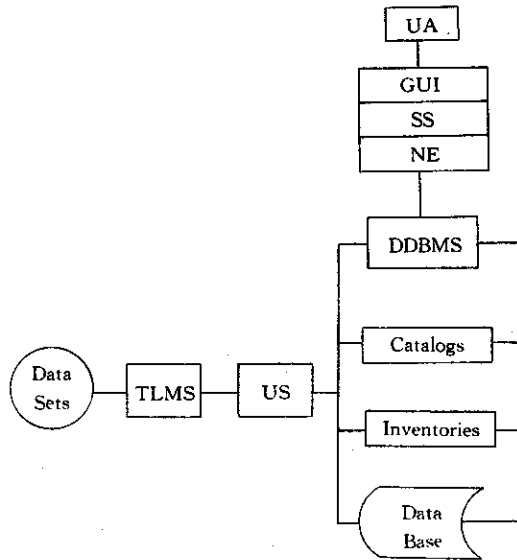


图 2 数据库与磁带库的集成结构

Fig. 2 The integrated structure for database and magnetic tape base

不入数据库的资料一律存放在磁带库数据集(Data Sets)中. 对标准格式磁带资料的检索, 其存取路径一般不把磁带中的数据装入数据库(Data Base), 从数据库中进行检索, 而是直接由磁带库管理系统(TLMS)实现. 但是, 用户可以从 DDBMS 和 TLMS 共管的客户服务(US)中进行查询, 通过目录(Catalogs)了解所需资料在何处, 知道需要访

问数据库还是磁带库。用户服务(User Service)是数据库与磁带库的接口。最终用户可以用数据库工具直接对数据库进行检索。

(6) 实时库、历史数据库和磁带库需要建立硬盘、光盘和磁带组成的多级存储系统(MLSS)。在当前和今后一定时期内,直接快速动态存取(与高速缓冲存储器联用)需要使用硬盘,作为一级存储系统。由可重写的光盘堆(Stacked Optical Disk File)或光盘自动换盘机组成的超大容量光盘存储系统,作为二级存储系统。大量数据集载体仍然是磁带,是联机的第三级存储系统。

(7) 各级数据库和磁带库所含的气候资料种类,根据 WWW 和 WCDP 的考虑,包括经由 GTS 和不经 GTS 传输的直接关系到大气-海洋-冰雪-陆地气候系统的资料有<sup>[2]</sup>:

① 高空资料(大气):气压、温度、风向、风速、湿度、水汽含量;

② 地面气候资料(大气-陆地):降水(雨和雪)、温度、气压、风向风速、蒸发、雪(覆盖、类型、深度、含水量)、水汽含量、日照、辐射、天气现象;

③ 海洋面和水下资料(海洋):海面风、温度、海表温度、海-气温差、海水温度与盐度廓线、海流、蒸发、降水;

④ 冰雪圈资料(海洋和陆地):冰川、大陆冰层、海冰边界、海冰覆盖、厚度、溶解与漂浮,雪覆盖和含水量;

⑤ 辐射收支(大气-陆地):有关的覆盖、类型、高度、厚度或光学厚度,行星辐射收支分量,太阳常数紫外线通量、地面反照率、地表辐射、红外辐射,陆地和冰面温度;

⑥ 大气成分资料(大气-陆地):CO<sub>2</sub>、O<sub>3</sub>和其它辐射性活动气体、甲烷、痕量气体,平流层 H<sub>2</sub>O 和气溶胶,对流层气溶胶、浑浊度、污染,大气与降水化学;

⑦ 水文资料(陆地):地面水(河、湖、水库;水流沉淀物输送/沉积,水的物理和化学性质及温度、冰覆盖的性质和范围),地下水(水位高度、温度);

⑧ 土壤与植被资料(陆地):蒸发/蒸散,植物水应力、地表及不同深度的温度、湿度,植物覆盖及变化。

(8) 气候资料分布式数据库,其内容主要包括如下:

① 气候资料要素库

地面基本资料数据库:国内地面气候资料数据库;全球地面气候资料数据库;全球地面气候月报资料数据库。

高空基本资料数据库:国内高空资料数据库(含探测);全球高空资料数据库(含探测);全球高空气候月报资料数据库。

农气报表资料数据库。

格点场资料数据库:地面、高空格点场历史资料数据库;模式输出物理量数据库。

辐射资料数据库。

其它要素库。

② 气候资料目录库

其内容主要包括数据库、磁带库、库存档案目录和国内外气象资料情报信息数据库。

③ 台站历史沿革库

其内容主要包括地理位置、仪器型号和记录时间、测量的有效性等,是数据库的元

数据(Meta data),是整个构成的重要组成部分。

### 3 气候资料分布式数据库运行环境

(1) 硬件 系统根据各级任务和规模,其主系统由客户/服务器网络体系结构组成。国家级要选用存储量大,吞吐能力强的高档服务器为基础;区域级要配置相应的中档服务器;省级要配置 PC 486 以上的服务器。外存储器国家和区域为硬盘—光盘—磁带三级系统,省级为硬盘—数据流(或光盘)二级系统。

(2) 软件 主服务器操作系统 UNIX,微机 DOS5.0 以上或 Windows/NT 以上。分布式数据库管理系统采用处理能力强的商品化数据库管理系统。

(3) 通讯 网络通讯环境主要考虑两个方面,一方面与气象通信系统(特别是国家气象卫星通信)的接口,包括 GTS/Internet;另一方面与国家公共数字数据网的接口。遵循标准的通讯协议(TCP/IP, OSI)。

(4) 本系统具有大规模数据库的特征 在一个数据中心必须要由多个数据库服务器组成,与此同时,系统必须有松耦合协同服务(Cooperative-Server)并行处理的能力。同时,在一个数据库服务器内部,也需要具有对称式(SMP)紧耦合(共享内存)多 CPU 结构。与此对应,数据库体系结构内部也必须有多线索(Multi-Thread)并行处理数据查询功能,以适应网络上多用户时检索的需要,提高应答服务效率。

### 4 气候资料分布式数据库服务系统

(1) 系统支持网络上任何节点进行存取,提供目录检索,包括浏览(Browse)、选择(Select)、查询(Query)等,以解决用户的下述问题:在什么地方存何种资料?怎样获取这些资料?感兴趣的资料质量如何?数量有多大?等等。

(2) 以交互方式对用户提供数据显示(Display),组合集成(Integration)和高层次分析应用,包括其支持软件 Graphics(图形)、Visualization(视算)、GIS(地理信息系统)等组合。在标准的图形用户接口环境下,提供各种应用服务。

(3) 系统对数据具有加工、更新、删除、插入、恢复、监控等功能,具有压缩(Pack)和解释(Unpack)的实用软件,用来维护管理资料和加工数据产品。

(4) 本系统可与国家的其它信息系统连接,透明的与异构数据库及文件互易操作,也可与其它国家信息系统和数据库有适当的联系。

### 5 系统建设实施策略

(1) 气候资料分布式数据库与网络通信极为密切,但系统建设可先行,不能等待通信网络建成后再起步,可分别建立国家、区域、省或地级的分布式数据库。随着通讯条件的成熟,以“滚雪球”的模式发展。

(2) 系统建设需要分步进行。数据格式规范化等可以先于数据库建设,各个数据库

也需根据条件有先有后, 建设分布式数据库是一个从局部到整体, 从低级到高级的发展过程。

(3) 国家-区域-省-地各级数据库的建设, 不必自下而上或自上而下, 可以上下同步, 但必须在总体统一规划下进行。

(4) 在近两三年内, 在现有的设备和商用 Sybase 数据库管理系统环境下进行分布式数据库建库和管理技术的研究试验。

国家气象中心气候应用室已于 1995 年建成中国地面气候资料分布式数据库。该库模拟分布处理环境并结合实际工作需要, 以 Sun 670 和一台微机构成双数据库服务器, Sun 工作站和若干微机作为数据库应用的客户机, 通过网络与有关的大型机相连。全国 670 多个地面基准气候站和基本气候站的逐日值和标准值及月值分别存放在两个服务器上。数据库由 Sybase 的数据库管理系统 SQL Server 和自行开发的软件进行管理, 达到数据检索的高效率和保证库中数据的一致性、完整性和安全性。该系统的建成为气候资料及时提供联机检索服务提供了保证, 提高了气候资料服务和气候资源开发的能力, 为今后建立包括多种报类的历史资料分布式数据库奠定基础, 也为本系统的建设提供经验。

### 参 考 文 献

- 1 李平安, 等. 分布式数据库系统概论. 北京: 科学出版社, 1992.
- 2 吴忠义, 马尚风. 世界气候资料计划文件汇集. 北京: 气象出版社, 1991.

## THE GENERAL FRAME OF DISTRIBUTED CLIMATOLOGICAL DATABASE

Hua Canhua

(National Meteorological Center, Beijing 100081)

### Abstract

The system goal, which is required to have unified standard in the respects of data format, data structure, use interface and user interface, is introduced in order to ensure data consistency and interchangeability. Thus the correlated systems of meteorological and non-meteorological departments can be received into a general frame which are connected and transferred with each other. And, the system structure is also described. It is a multiple data structure system with the geographically state-regional-provincial-prefectural levels climate database, but under the control of distributed database managing system. Logically, it is an entire database.

**Key words:** Climatological Distributed database System goal and structure