

# 气候资料新一代信息存储与服务系统的研究

郭发辉

(国家气象中心,北京 100081)

## 提 要

着重阐述了气候资料新一代信息存储与服务系统的技术基础,主要有数据库技术、数据仓库技术、数据开采技术、环球网技术和智能代理技术等,同时也阐述了系统的结构、功能以及实现的策略。

**关键词:** 气候资料 新一代信息存储与服务系统 系统的结构和功能

## 引 言

气候信息资源是我国信息资源的重要组成部分。随着气象和气候科学在指导人类社会发展中作用的提高,社会各行各业对气候资料的需求越来越迫切,其要求也越来越高。因此,气象资料工作必须面向市场、面向社会,加强气候信息资源的开发和利用。

气候资料工作的宗旨是服务,而高效率的服务需要有先进的信息存储、检索技术基础。多年来,为适应气象业务发展的需要,国家气象中心在气候资料现代化建设和数据开

发方面做了大量的工作,气象资料的信息存储设备已有了较大的规模,气象资料服务有了较大的提高。但是,随着信息技术的迅猛发展,人们的工作方式、相互关系、思维观念等诸多方面将发生变革,气象资料服务的方式和规模也应有相应的变化。因此,如何使气象资料服务工作更好地适应国民经济的需要,如何面向市场提供气候环境信息,都是值得我们重视的问题。

根据国情,我国信息化建设与国外的信息高速公路的内涵有所不同,重点强调信息

化体系中的信息资源、国家信息网络、信息技术应用、信息技术与产业、信息化人才、信息化政策法规和标准等六个要素之间的紧密关系。而且其中信息资源的开发利用将放在核心地位,突出建立自主的信息产业。这也就迫切需要在抓好气候资料的基础业务建设的同时,加强对利用高科技建立新一代气候资料信息存储与服务系统的深入研究。为此,通过对国内外气候资料信息方面的了解和信息技术的研究,结合我国的具体情况,提出了新一代气候资料信息存储与服务系统的框架。

### 1 气候资料新一代信息存储与服务系统的技术基础

我们传统的信息存储与服务系统主要是存储和提供通用性的资料产品,所用的技术一般是基于文件管理和数据库管理。新一代信息存储与服务系统则要求面向市场,也就是说要适应大量多变的客户化、定制化应用,将信息服务产业化,向用户提供有高科技含量的信息产品,这在技术上必须有一个根本的转变。

20世纪90年代以来,计算机新技术迅速发展,作为气候资料新一代信息存储与服务系统,以下五项应成为它的技术基础而加以重视和深入探讨。

#### 1.1 数据库(Database)技术

数据库技术从1968年美国IBM研制了世界第一个数据库管理系统IMS<sup>[1]</sup>至今,只不过30多年的历史,但它已成为当今计算机系统有效利用信息的核心和基础,也是一项成熟的技术。

目前流行的数据管理技术虽仍是关系数据库管理技术,但它在管理磁盘中的数据基础上,出现了关系对象数据库管理技术。它扩充了管理多媒体数据的能力,可在关系数据库的表中增加能支持一个或多个的多媒体字段(如大文本、字段、图象字段等)的功能。

同时,关系对象数据库采用外连文件技术,即多媒体数据不直接写入数据库表中,而是将多媒体数据形成的文件索引写入表中,数据本身仍以文件形式存放在磁盘中,接受来自数据库管理系统所发的命令的操作。此

外,还可为用户提供用户自定义数据类型和操作函数应用程序接口的支持技术。

#### 1.2 数据仓库(Data Warehouse)技术

在气候资料工作中应用数据库技术,由于有良好的查询优化机制和处理过程开发工具,比直接在文件基础上进行程序加工无疑是一大进步。但是,要进行专业化的数据分析,涉及多维数据视图的概念,仅靠关系数据库就力不从心了<sup>[2]</sup>。维是人们观察现实世界的角度,不同专业需要从不同的角度去观察分析数据,这就必须把数据从关系数据库中抽出,按专业需求物理地重组。这种能够灵活地重新组织对用户呈现出多维数据视图的概念就是数据仓库(DW)。

随着服务器/客户机技术日渐成熟和并行数据库的发展,数据仓库技术不断得到发展并趋于成熟。数据仓库是把一个单位的历史数据收集到一个中央仓库以便于处理,它是面向应用(主题)的、集成的、支持决策过程的、随时间而变的、持久的数据集合,是当今信息管理的主流趋势。

气候资料数据库在数据预处理的基础上加载,已具有一般的数据一致性,但并不是对任何应用都是完全协调一致的,而数据仓库对相关主题来说则必须是一个完全协调一致的信息体。数据库在基层,数据仓库在上层。它是近代数据库技术发展联机分析OLAP的前提。从传统的数据库应用环境向以数据仓库为基础的应用环境转移,是近代信息服务系统发展的必然趋势。

#### 1.3 数据开采(Data Mining)技术

数据开采(DM)是近代数据库技术面对复杂应用而发展起来的新技术,也可以认为它是关系数据库联机分析的智能扩展<sup>[3]</sup>。数据开采正好与数据查询相反。数据查询的问题和答案都是肯定的,而数据开采是面对不确定的提问,回答隐藏在数据背后的未知东西。1991年Inmon提出数据仓库的概念,1993年关系数据库创始人Codd就提出联机分析概念,可见大家认识到仅有数据库是不够的,还必须要有数据仓库,更需要分析工具。数据开采涉及数据库及人工智能等多学科技术。由于专家系统存在着知识获取的瓶

颈现象,所以人们寄希望于在数据库支持下,从大量事实中通过机器学习来完成知识的自动获取,也就是知识发现。人工智能的知识发现是指从数据库发现有用的知识的过程,数据开采是知识发现过程中的一个特定步骤。它是一个反复渐进的过程,通常包括数据清理、提出假设、选定算法、反复匹配、验证规则、评价解释、构成知识、应用输出。方法中包含可视性(Visualization),即可以透过数据观测不同层面的细节,发现并解释客体与数据之间的相关性,并以图形显示的直观方式表达。

气候数据库是一个知识的海洋,怎样发现和提取隐藏在其中的知识,如何展现历史事实,特别是各种商业活动与气候历史之间的相关特征,这就要应用 DM 技术。这是非常有经济价值和广阔前景的信息服务。如果说网络服务提供者(ISP)是信息高速公路上跑的车,那么应用 DM 技术的信息服务就是车上装的货——信息内容提供者(ICP)之一。现在人们已认识到 ICP 和联机服务正在成为制约我国信息化建设进一步发展的重要因素。

#### 1.4 环球网(World Wide Web)技术

环球网(WWW)技术是1990年欧洲量子物理实验室 CERN 为统一管理各种资源,交流研究成果而发展起来的网络技术。它是由传输控制协议/公共网络协议(TCP/IP),超文本传输协议(HTTP),超文本标记语言(HTML),统一资源地址(URL),浏览器(Browser),网点服务器(Web server)等这些基本概念构成<sup>[4]</sup>。不但对多媒体信息具有强大的统一管理功能,而且能作双向沟通。当前,WWW 技术已经成为 Internet 的主流。

WWW 以超文本标记语法描述各种多媒体文件,提供 Web server 上的各种信息都是超链接的,容易使用 Browser 方式呈现在用户面前。Web server 是 WWW 技术的核心,HTTP 的服务器和浏览器之间采用的传输协议,它的底层是 TCP/IP,任何形式的数据包括文本、图象、声音、影视均可相互传输交换。随着 Internet,特别是 Intranet(内部用户网)和 Extranet(外部特定用户网)的兴

起,出现了一类新兴人才—Webmaster(Web 设计管理者),他们掌握信息系统的核心技术。

从原来的客户/服务器(Client/Server)网络系统向 WWW 转移是基础设施上的逻辑覆盖,在软硬件上要作适当的配制调整。例如,原有的 C/S 网络系统(Sun670-Wintel PCs),只在服务器高端装上 CERN httpd (UNIX 环境下的 Web server 免费软件),在 PCs 低端装上 Netscape(Windows 环境下的 Browser 软件),即可形成一个简单的 Intranet。在 PC 机浏览器窗口上指定起止日期和时间,就可以启动服务器中的 HTML 文本,其中嵌入有 Java Applet 应用程序,即能动态显示云图或天气现象(如降水过程)的演变情况。

#### 1.5 智能代理(Intelligent Agent)技术

自1994年以来,智能代理(IA)和 Internet 一样成为热门话题。因为在广泛使用 C/S 模式的环境下,系统规模迅速扩大,运行任务日渐繁重,维护管理越趋复杂,严重影响工作效率。为解决这个问题,IA 应运而生<sup>[5]</sup>。

IA 机理是能接受应用委托,根据应用程序和数据库管理的需要,随机组合自动执行指定任务,而用户不必关心应用程序、数据库在网络上处于什么位置和什么状态。也就是说,IA 具有以下特性:

(1)推理性:用户把任务交付后,IA 能根据所需要的资源自动地作出判断完成作业。

(2)学习性:能在执行中记忆过去运行的状态,不断增加自身对各种事件的处理能力。

(3)协同性:能根据系统资源状态,对系统承接的任务之间进行动态调整。

(4)移动性:在协调多项任务时,能从一个系统移至另一个系统,甚至可选择本地或远程去执行。

(5)自主性:任务委托后,可以完全脱离用户,自动执行,可根据某一事件(如某项任务完成),或指定时间触发执行。

IA 技术反映了当代多服务器集中管理所具有的高性能计算机特征。如果气候资料信息服务系统要面向市场,快速响应复杂多

变的各种应用,没有多服务器聚合并行处理集中管理的 IA 能力是不可想象的。IA 将成为气候资料新一代存储与服务系统的控制中枢。

## 2 气候资料新一代信息存储与服务系统的结构和功能

在近代计算机新技术的强劲推动下,信息系统的体系结构经历了集中—分布—再集中螺旋式的跃进过程,即从集中式的主机(Main frame)到分布式的客户/服务器(Client/Server),又很快演变到在分布基础上再

集中的 BWD (Browser—Web server—Database)模式。

BWD 体系结构以 Web server 为中心,采用 TCP/IP, HTTP 等标准协议,用户通过 Browser 访问 Web 及其后台的 Database。气候资料新一代信息存储与服务系统也必然要逐步从 C/S 转向 BWD 才能适应信息服务面向市场的客观需要。

从总体结构看,新一代信息存储与服务系统应该由 5 个平台构成(见图 1)。

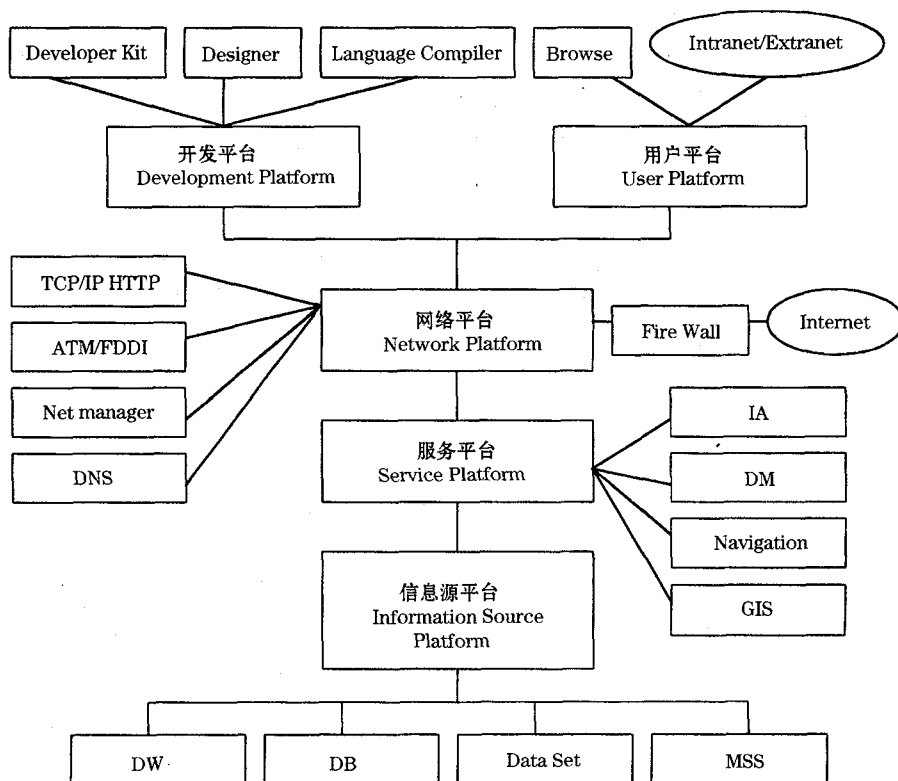


图 1 系统总体结构图

(1)网络平台:主要功能是网络监控和网络管理,包括域名管理系统(DNS)的职能,通过防火墙与 Internet 相联。

(2)服务平台:主要功能是查询导航,数据开采,由智能 Agent 集中管理,其中包括有地理信息系统(GIS)空间信息管理网上查询的能力。

(3)开发平台:具有各种开发工具,设计

套件、语言编译等软件,是一个开发集成的支撑环境。

(4)用户平台:运用 Browser,对单位内部包括国家气象中心(NMC)、国家气候中心(NCC)及外部特定用户提供服务窗口。

(5)信息源平台:统一和高效管理气候资料信息,包括数据库数据和数据文件数据集。软件的层次结构见图 2。最底层是信息

源。一个完整的对全国综合性适用价值的气候环境多维信息源,最基本应该具有:(1)逐日定时全国地面要素分布(Text);(2)逐日定时全国云系(Image);(3)逐日定时北半球环流场(Grid)等。在基准数据集之上有文件管理系统和数据库管理系统,再上面有DW控制部件和DM分析工具,最上层有Agent智能管理各种客户化应用程序,提供快速反应的联机(On-Line)服务。

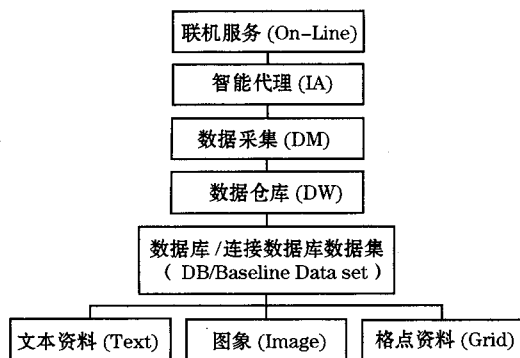


图2 软件层次结构图

图3是系统硬件配置图。网络的性能通常以服务器上的访问信息的容量和速度来评价。如果服务器与应用访问只有直接的一层交换,那么在查询频率不高的情况下尚能运行。如果大量应用需要分布在多台服务器上,这种环境需要建立第二层交换,即服务器交换。新一代信息存储与服务系统需要面向市场的响应能力,必须建立这种硬件环境。因此,它的核心设备是具有多CPU并行处理能力的多台64位高性能聚合服务器作为群

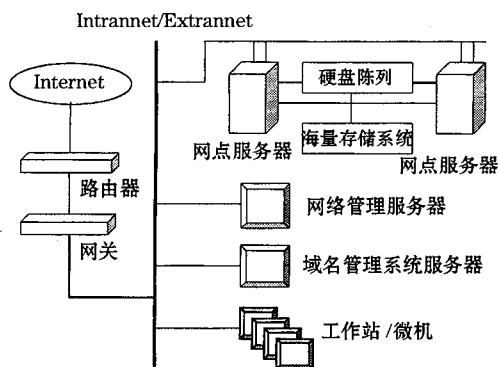


图3 硬件配置结构图

Web server,并在GB级硬盘阵列(Disk Array)和TB级海量存储系统(Mss)的支持下工作。

气候资料新一代信息存储与服务系统的基础是创建完全协调一致的信息源,系统的核心是配备高性能计算机服务器Web server,系统的关键是具有智能代理的先进软件。

总起来说,气候资料新一代信息存储与服务系统将具有以下五大功能:

(1)在IA的支持下,能面向气候变化、防灾减灾和商业活动开展快速反应的WWW在线服务;

(2)在DM系统的支持下,不仅能提供常规气候数据加工服务,而且能提供对客户需求的有针对性的多样化定制化信息导航服务;

(3)在GIS的支持下,能使气候数据与相关地理信息融合,提供统一管理空间信息的图形检索可视化服务;

(4)在开发平台环境的支持下,能够提供应用软件的高效开发和优化集成的服务;

(5)在MSS的支持下,能够高速调度气候基准数据仓库,再现和分析气候历史事实,为气候资源的开发利用提供综合气候信息服务。

### 3 实现策略

(1)改变传统提方案一定计划一实施的做法,而应实行试验一议方案一定计划的新做法,即先试验,从中得出结论,以此提出可行方案,再定逐步实施的计划。只有这样,才能跟上技术的迅速发展。

(2)气候资料信息存储与服务系统是一个多学科、多技术交叉的复杂的系统工程。因此,不一定要一步到位,而是在现有设备基础上,利用BWD体系结构,在Intranet上试行一个特定项目,即可把目前的一些常规资料产品服务逐步从C/S转向体系结构,以此作为突破口,取得实效,由点到面,由内转外,面向市场,逐步形成开放型的气候资料信息存储、服务系统。

(3)面向市场开展信息服务是一个商业行为,风险和收益同在,要先找到切入点,才能逐步展开,把国家投资与自身发展结合起

来,起步时可与 ISP 共同利用资本运营手段,达到资源优化,优势互补。

#### 4 结束语

本文所提到的技术基础,都是世界当前成熟的先进的信息技术,所需要的系统软件都有商业产品,国内有关高校、科研单位也都在开发有自主知识产权的实用产品,并且许多部委也发表了类似的创建 Intranet 成功经验的报道。我们在试用 WWW 技术方面也有一定的实践经验,说明是可行的。该方案不仅适用于国家级气候资料部门,而且也适用于省级气候资料部门。省级气候资料部门仅根据自身的需要在内容和规模上做些调整就

行。方案的实现将使我国气候资料工作步入世界先进水平,但它的实施是一个十分复杂的过程,需要多方面的配合和协作。

#### 参考文献

- 1 冯玉才. 数据库系统基础. 武汉:华中工学院出版社, 1984.4.
- 2 王珊,罗立. 从数据库到数据仓库. 计算机世界, 1996. 7.15.
- 3 陈文伟,邓苏,张维明. 数据开采与发现综述. 计算机世界, 1997.6.30.
- 4 刘锦楠,万云龙. WWW 文件设计——HTML 语言实现,北京:机械工业出版社,1997.
- 5 杨福泉. 网络中的软件代理,计算机世界, 1996.10.28.

## General Frame Research for New Generation Information Storage and Service System of Climate Data

Guo Fahui

(National Meteorological Center, Beijing 100081)

#### Abstract

Technical fundamentals of the new generation information storage and service system of climate data, including database, data warehouse, data mining, world wide web and intelligent agent techniques are given. And the structure and functions of the system, as well as system implemental strategy are also described.

**Key Words:** climate data new generation information storage and service system structure and function of the system