

新一代气候资料信息服务 系统框架的探讨

花灿华 盛永宽 许 松

(国家气象中心,北京 100081)

提 要

着重阐述了新一代气候资料信息服务系统的技术基础,主要有数据仓库技术、数据开采技术、环球网技术和智能代理技术。同时也阐述了系统的结构和功能以及实现的策略。

关键词: 气候资料 新一代信息服务系统 系统结构和功能

引 言

气候资料工作在气象部门是一项基础性的业务工作。多年来,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,取得了明显的成绩,大大提高了气候资料基础业务自动化的程度,也为开展快捷、高效、准确、全方位的气候资料服务打下了坚实的基础,相应地提高了气候资料服务的效益。

近年来,信息技术进步很快,如何利用新信息技术,发展我国的气候资料服务系统是我们所面临的问题。我国的信息化建设与国外的信息高速公路的内涵有所不同。发达国家如美国强调建立千兆级结点(gigapop)之间超高速、超带宽相互连通的联合体。我国根据国情,则强调信息化体系6个要素(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)之间的紧密关系。我们通过对国内外气候资料信息方面的了解和对信息技术的研究,结合我国的具体情况,提出了新一代气候资料信息服务系统的框架,为下一步具体进行

技术设计提供依据。

1 新一代气候资料信息服务系统的技术基础

传统的气候资料信息服务系统主要是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。新一代气候资料信息服务系统则要面向市场,也就是说适应大量多变的客户化定制化应用,要把我们的信息服务产业化,提供有高技术含量的信息产品,这在技术上必须有一个根本的转变。

近代特别是90年代,计算机新技术更新很快,其中值得我们高度重视和深入探讨的有以下四项。

1.1 数据仓库技术(Data Warehouse)

在气候资料工作中应用数据库技术,极大地提高了检索功能和服务效率,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^[1]。例如,关系库对多点多要素长序列加工统计比较高效,但要作平面(场)的多要素相关分析就困难了,维是人们观察现实世界的角度,不

同专业需要从不同角度去观察分析数据,这就必须把数据从关系库中抽出,按专业需求重组。这种能够灵活的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)。因此,在本质上数据库是面向处理(操作)的,而数据仓库是面向应用(主题)的。另一方面,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信息服务驱动的以数据仓库为基础的应用环境转移,是近代信息服务系统的必然趋势。组成数据仓库应用环境的基本部件应该有:①定义部件(设计和构造数据仓库的执行系统),②采集部件(将源文件数据或数据库数据转换到数据仓库),③管理部件(用于数据仓库的管理工作,包括安全、备份、监测等),④目录部件(提供有关数据仓库数据的存贮位置、类别属性等信息,包括元数据),⑤分析部件(提供有关数据访问和分析的工具)。

1.2 数据开采技术(Data Mining)

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库联机分析的智能扩展^[2]。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,而数据开采则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不够,还要数据仓库,更要分析工具。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知识的整个过程,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。方法中包含可视化(Visualization),即可以透过数据观察不同层面的细节,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。

怎样发现和提取隐藏在气候数据库的知识,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,这就要运用 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,那么使用 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的重要因素。

1.3 环球网技术(World Wide Web)

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(Webserver)等基本概念构成^[3]。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。发展到现在,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。在 WWW 世界中 Browser 扮演重要角色。现在许多 Browser 增加了对 Java(一种面向对象跨平台人们称之为世界语的编程语言)的支持,使目前主要用的 HTML 语言由只能静态表示变成动态表演的新世

界。Web server 是 WWW 技术的核心, HTTP 是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、图像、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)。

从原来的客户/服务器(Client/server)网络系统向 WWW 转移和 Extranet 均是基础设施上的逻辑覆盖,只需根据规模需求,在软硬件上作适当的配置调整。例如,我们原有 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,就可以启动服务器上的 HTML 文本,其中嵌有 Java Applet 应用程序,即能动态显示云图或天气现象(降水过程)的演变情况。

WWW 是信息服务系统开展联机服务(On-Line)的基础技术,这是高新技术给我们的挑战和机遇。如果说我们的信息服务系统运用文件系统和数据库技术比先进国家落后 15—20 年,那么只要抓住新的机遇,我们就可以与先进国家几乎在同一起跑线上。

1.4 智能代理技术(Intelligent Agent)

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^[4]。

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,而用户不必关应用程序、数据库在网络上处于什么位置什么状态。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,甚至可选择本地还是远程去执行。

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(如什么任务完成),或指定时间触发执行。

智能代理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 将成为新一代气候资料服务系统的控制中枢。

2 新一代气候资料信息服务系统的结构和功能

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)到分布式的客户机/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。新一代气候资料信息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,新一代气候资料信息服务系统应该由五个平台构成(见图 1)。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)的责能,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,是一个开发集成的持撑环境。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。

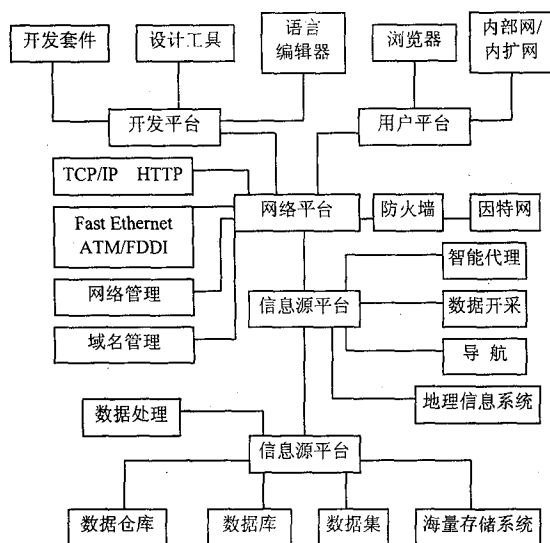


图1 新一代气候资料信息服务系统构造框图

硬件的配置结构(图3)。服务器上的访问信息的容量和速度是影响网络性能的因素之一。如果服务器与应用访问只有直接的一层交换,则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群作Web server,并在GB级硬盘阵列(Disk Array)和TB级海量存贮系统(MSS)的支持下工作。

软件的层次结构(见图2)。最底层是信息源。一个完整的对全国有综合性实用价值的气候环境多维信息源,最基本的应该具有: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服务(On-Line)。

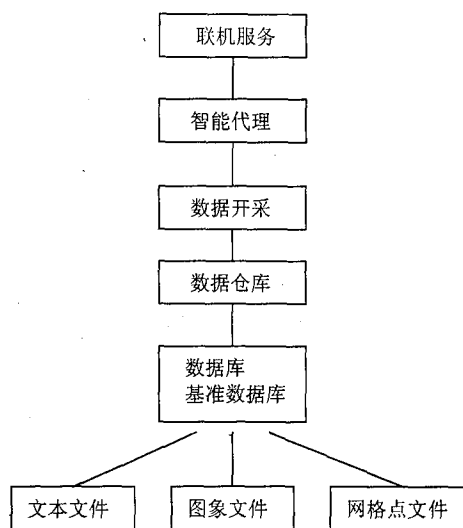


图2 新一代气候资料信息服务系统软件层次结构

气候资料新一代信息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服务系统将具有以下五大功能:①在Agent智能管理的支持下,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集,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。

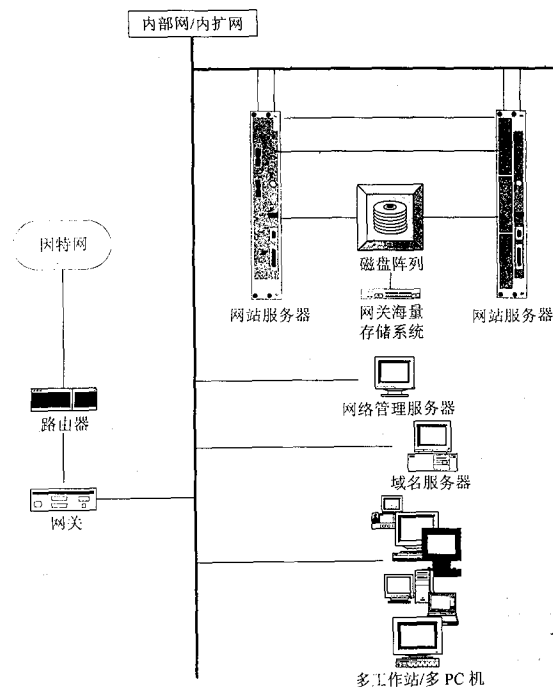


图3 新一代气候资料信息服务系统硬件配置结构

3 实现策略

①改变传统的做法,提方案一定计划一再实施,而应实行先试验—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,才能跟上迅速的发展。

②新一代气候资料信息服务系统是一个多学科、多技术交叉的复杂系统工程。因此,不要一步到位,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 BWD 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,要把国家投资与自我发展结合起来,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,增值分享。

4 结束语

本方案所提及到的技术基础,都是世界公认的当代先进的成熟信息技术,所需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 成功经验的报导。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资料部门,而且也适用于省级资料部门。省级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。本方案的实施是一个十分复杂的过程,它需要多方面人员的配合和协作,其中尤其是领导的重视和参与更为重要,这些因素是该方案顺利实施的重要保证。本方案中所提到的网络技术的应用,必须在气象部门的统一规划下进行。本方案的实现将使我国气候资料工作处于世界领先水平。

参考文献

- 1 王珊,罗立.从数据库到数据仓库,北京:计算机世界(CHINA COMPUTERWORLD),1996.7.15.
- 2 陈文伟,邓苏,张维明.数据开采与知识发现综述,北京:计算机世界(CHINA COMPUTERWORLD),1997.6.30.
- 3 刘锦楠,万云龙.WWW 文件设计—HTML 语言实务,北京:机械工业出版社,1997.1.
- 4 杨福泉.网络中的软件代理,北京:计算机世界(CHINA COMPUTERWORLD),1996.10.28.

(下转 14 页)

General Frame Approach for New Generation Information Service System of Climate Data

Hua Canhua Sheng Yongkuan Xu Song

(National Meteorological Center, Beijing 100081)

Abstract

The technical fundamentals of the new generation information service system of climate data are described, which consists of data warehouse technique, data mining technique, world wide web technique and intelligent agent technique. The structure and function of the system, as well as implemental system strategy are presented.

Key words: climate data New generation information service system system structure and function