

文章编号: 1000-2022(2000)-01-0139-06

## 气候资料信息处理系统展望\*

许 松, 花灿华

(国家气象中心, 北京 100081)

**摘要:**着重阐述未来气候资料信息处理系统的技术基础, 主要有系统集成、可视化技术、数据仓库技术、数据开采技术、环球网技术和智能代理技术。同时也阐述了未来系统的结构和功能。

**关键词:**气候资料; 信息处理系统; 系统结构和功能

**中图分类号:** P468.0      **文献标识码:** A

气候资料工作在气象部门是一项基础性的业务工作。它既是经济建设、国防建设和科学研究不可缺少的重要依据, 又是搞好气象服务和气象科研工作不可缺少的基础, 并随着气象和气候科学在指导人类社会发展中作用的变化而日益显示出它的重要性。为了开发利用气候资料资源, 充分发挥气候资料的潜力, 以适应现代化建设和气象业务深入发展的需要。必须加强对利用先进的信息技术建立新型的数据处理业务系统的研究, 提高气候资料数据自动化的水平和气候资料应用服务的能力。这是气候资料工作当前迫切需要解决的一个重要的研究课题。

### 1 未来信息处理业务系统的技术基础

90 年代, 计算机新技术不断推出。作为未来气候资料处理业务系统的技术基础, 有以下 6 项值得深入探讨。

#### 1.1 集成技术(System Integration)

所谓系统集成就是充分利用计算机软件、硬件、网络、数据库等先进技术, 以计算机网络为依托, 以新型计算机模型为基础, 以分布式数据库为中心, 以先进的软件支撑环境, 实现分布处理, 统一控制和管理。回顾系统集成的发展过程, 大致经历了三个阶段。这三个阶段是相互联系、相互渗透的。第一阶段是 70 年代, 系统集成的核心是文件管理系统, 多个应用程序集成在一起, 文件管理系统统一管理程序和数据。第二阶段是 80 年代, 系统集成的核心是数据库管理系统, 通过公共的数据库, 将信息系统统一管理起来。第三阶段是 80 年代末 90 年代初, 系统集成是基于网络计算模式, 或者说是基于 Intranet 的新一代管理信息系统(MIS)。网络计算模式以 Web 为中心, 采用 TCP/IP, HTTP 为标准协议, 客户端通过 Browser 访问 Web 以及 Web 相连的后台 Database, 所以也称为 BWD 模式。它突破了传统的 Client/Server 模式, 是一种多层次的 Client/Server 模式, 使信息系统之间由于系统结构、网络通讯、文档格式、文件标准等

\* 收稿日期: 1999-02-08; 改回日期: 1999-07-12

作者简介: 许 松, 男, 江苏兴化人, 1950 年 5 月生, 高级工程师。主要研究方向: 气候资料处理业务系统

环节不统一的分离局面,通过 Client/Server 的标准技术,以方便的低成本的方式达到无缝连接的统一整体。这个阶段的系统集成特点是采用框架( framework) 软件进行统一管理,集有效数据管理和良好过程控制于一身。它提供了一个集成平台(BWD 模式逻辑结构如图 1)。

为了进一步提高气候资料业务系统的自动化水平和应用服务的服务能力,未来气候资料信息处理业务系统必须采用 BWD 模式,将气候资料数据处理、数据管理和应用服务集于一体,提供一个集成平台,以便实现分布处理、统一控制和管理,达到高效利用计算机资源,提高整体的业务系统的效益。

## 1.2 可视化技术( Visualization)

所谓可视化技术,就是在做数据分析时,用交互式、动态模拟、图形表示。即做到如同俗话说“看得见,摸得着”。在数据分析过程中,使用者只与图形打交道。即:通过在屏幕上作图,并对一些异常情况作动态模拟修改,修改后的图形再自动生成数据文件。

可视化技术与一般计算机绘图,以及图形化或 CAD 软件在编程指导思想和方法上有所不同。其一,数据分析可视软件的主体是数据分析本身,而图形及图象是表现形式之一。它使人们能深入到数据分析过程中,不仅能看得见,还能掌握其分析过程及相互关系,并及时作出反映。其二,强调高度的智能性。正是由于以上原因,开发此类软件不仅需要清理分析模块当前各部分的关系,还要建立合理可靠的各阶段的联系,让与它有关的方方面面、前前后后都能及时作出反应并修改。就是所谓的“牵一发而动全身”。其三,强调图形与数据迅速频繁地进行实时交换。其四,以使用鼠标器的屏幕操作为主辅以少量的键入。因此,开发此类软件应有功能极强的图形编辑手段。可视化技术在气象部门已得到应用。德国气象局开发了一个气象数据可视化软件 TRITON,目前已应用于日常的天气预报中。国家气象中心与清华大学合作开发的“三维气象动态图象系统”,可以清晰地看到气象物理量的分布特征和演变特点,能十分直观地了解大气的三维结构。同样,应用可视化技术进行气候资料信息处理,可以加快数据处理的速度,提高气候资料处理质量,提供形象、直观的气候资料服务,提高气候资料服务的社会效益和经济效益。

## 1.3 数据仓库技术(Data Warehouse)

在气候资料工作中应用数据库技术,已极大地提高了检索功能和服务效率,因为它有良好的查询优化机制和处理过程开发工具。一般来说,它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>[1]</sup>。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,有时甚至不如文件系统。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,按专业需求物理地重组。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。另外,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的前提。从传统的数据库应用环境向信息服务驱动的以

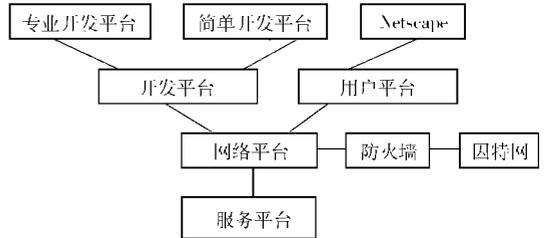


图 1 BWD 模式逻辑结构

Fig. 1 Logical structure of BWD model

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。

#### 1.4 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<sup>[2]</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,事隔2年,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。

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,那么我们 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。

#### 1.5 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>[3]</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。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 语言原先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 转移和 Extranet 均是基础设施上的逻辑覆盖,并不要抛弃原有设施新建物理网络,只要根据规模需求,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。

### 1.6 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<sup>[4]</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 未来气候资料信息处理业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。气候资料处理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,未来气候资料信息处理系统应该由六个平台构成(见图 2)。

(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)数据处理平台:主要功能是对气候资料数据进行格式检查和质量控制,以保证数据的准确性。

(3)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。

(4)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。

(5)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。

(6)数据管理平台: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。

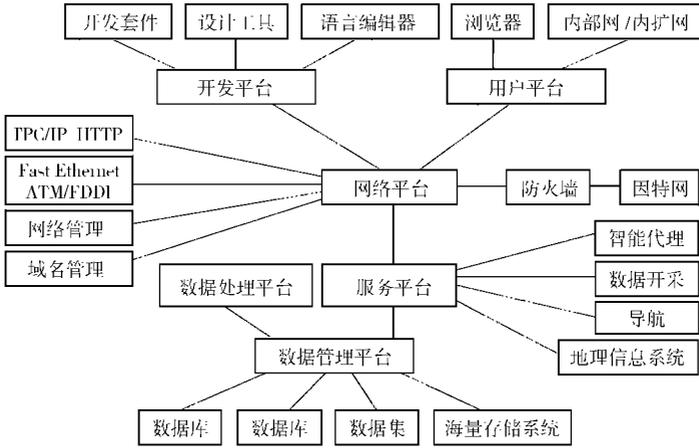


图 2 未来气候资料信息处理系统

Fig. 2 Climate data processing system in the future

软件的层次结构见图 3。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) 陆地、海洋观测和统计分析; (2) 逐日定时全国云系 (Image); (3) 逐日定时北半球环流场 (Grid); (4) 数值预报、卫星定量、雷达产品等。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)。

硬件的配置结构见图 4。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。未来信息处理业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 级海量存储系统 (MSS) 的支持下工作。

气候资料未来信息处理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

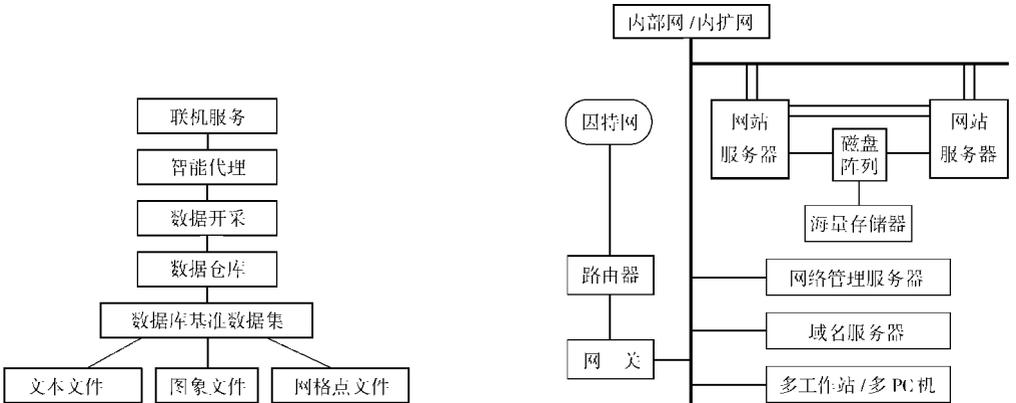


图 3 软件的层次结构

Fig. 3 Structure of software design at different levels

图 4 信息服务系统硬件配置结构

Fig. 4 Structure of hardware configuration of data service system

性能计算机 Web server, 系统的关键是具有智能代理的先进软件。

总的来说, 气候资料信息处理系统将具有以下六大功能: (1) 在可视化数据分析软件的支持下, 实现可视化气候资料数据的格式检查和数据的质量检验; (2) 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; (3) 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; (4) 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; (5) 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; (6) 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。

### 3 结 语

(1) 未来气候信息处理系统是一个多学科、多技术交叉的复杂系统工程。因此, 不要求一步到位, 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, 可把目前的一些信息处理软件逐步从 C/S 转向 BWD 体系结构, 可以此作为突破口, 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象信息服务系统。

(2) 上述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 的成功经验。我们在试用 WWW 技术方面也有一定的实践经验, 这说明我们所提出的未来气候信息处理系统的设想是符合未来发展趋势的、是可行的。

(3) 未来气候信息处理系统的建立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。未来气候信息处理系统的实现将促进我国气象工作向世界领先水平迈进。

### 参考文献:

[1] 刘锦楠, 万云龙. WWW 文件设计——HTML 语言实务[M]. 北京: 机械工业出版社, 1997

## Prospect of climate data processing system

XU Song, HUA Can-hua

(National Meteorological Center, Beijing 100081)

**Abstract:** Described mainly are technical fundamentals of future climate data processing system, including techniques of system integration, visualization, data warehouse, data mining, world wide web and intelligent agent. And the structure and function of the system are also discussed.

**Key words:** climate data; information processing system; system structure and function