

气象数据文件快速下载服务系统的设计与实现^{*1}

高峰¹⁾ 王国复¹⁾ 喻雯¹⁾²⁾ 冯明农¹⁾ 孙超¹⁾ 罗琦³⁾

¹⁾(国家气象信息中心,北京 100081) ²⁾(北京航空航天大学,北京 100191)

³⁾(甘肃省气象信息与技术装备保障中心,兰州 730020)

摘 要

有效地为专业和业务用户提供快速批量下载气象数据服务,需要一种新型数据服务系统,以弥补现有气象数据共享服务系统的不足。该文在分析气象资料服务系统发展阶段和技术特点的基础上,根据业务需求设计了一个基于元数据技术的、针对专业和业务用户的、能快速批量下载数据的新气象资料服务系统。该系统基于前台应用系统和后台管理系统实现了批量下载、文件目录导航、数据同步更新、多线程断点续传、安全管理和灵活配置等功能。该系统是对气象资料服务方式的一种全新设计,系统发布后得到专业用户的肯定,具有广泛的推广价值,对气象数据服务系统的技术开发有较好的参考作用。

关键词: 数据服务; 元数据; 目录浏览; 增量同步

引 言

气象资料工作的主要环节包括:采集(观测)、传输、加工处理、存储、服务和归档等。其中,服务是直接面对用户的一个环节,也是唯一由资料生产者和管理者以外人员参与的环节,服务的好坏由满足用户需求的程度决定。服务有在线服务和离线服务两种形式,在线服务通过气象资料服务系统来实现。气象部门较早就开始了气象资料服务系统的开发与建设,其发展历程大体可分为 4 个阶段^[1-4]:①单机文件检索阶段。20 世纪 80 年代初,随着地面气象资料信息化工作的展开,气象资料管理中逐渐引入各类计算机,初步改变了长期以来纸质报表存储数据和人工抄录服务的资料存储检索工作模式。省级气象部门主要采用 CS400 微型计算机来存储信息化的气象资料并基于文件系统为内部业务科研用户提供检索服务;国家级气象部门采用 FM360 大型计算机,依靠机器配套的磁带机来进行数据存储和检索。在这一阶段,是以信息化处理为主,数据的检索服务还未作为工作重点,所以尚未形成专门的数据检索服务系统,个别的数据检索服务也基本上在单机上运行。②接口调用阶段。从 20 世纪 90 年代中期开始,以实时气象资料数据库的建成投入运行

为标志,气象部门的数据服务迈入了局域网络管理阶段,特别是“八五”和“九五”期间,通过大型国家基建项目支持,国家级和各省市气象部门均建成了主要面向业务的各类数据服务系统(专用的文件库),形成了专门承担数据存储和检索功能的服务器,内部局域网络上的用户均可以访问并通过特制的接口来调用所需数据。③数据库系统阶段。到 20 世纪 90 年代末期,以“9210 工程”投入运行为标志,逐步形成了国家、区域、省、地市 4 级分布式实时数据管理系统。以此为契机,许多省级气象部门开发了为天气预报等业务提供服务的实时资料查询系统,也有部分省级气象部门开发了以有偿服务、气候分析为目的的气候资料查询系统等。这个阶段开发的资料服务系统主要特点是:采用商用数据库管理系统(RDBMS),提供单一的数据查询功能,无统一的数据规范,未采用元数据技术,大量的历史气象资料管理没有与之相适应的大型综合存储检索系统支持。④分布式数据服务阶段。从 21 世纪初开始,随着国家信息化进程的加快,气象部门也日益注重数据的科学管理和高效利用,“十五”期间先后实施国家级气象资料存储检索系统、气象科学数据共享系统的建设,旨在集约化和标准化原则下,充分利用数据库和元数据技术,构筑海量、分布式、网络化的综合数据服务系统。这个阶段系统建设以数据为中心,以顶层设计为前提,以元数

* 国家科技基础条件平台工作专项资金项目(2005DKA31700)资助。
2009-07-21 收到,2010-02-09 收到再改稿。

据技术为核心,以数据标准规范为基础。

气象资料覆盖了地球五大圈层:大气圈、水圈、岩石圈、冰雪圈和生物圈,应用领域广泛,这就决定了气象资料服务系统的复杂性^[2]。目前运行的大型气象资料服务系统包括国家级气象资料存储检索系统、气象科学数据共享平台等,这些系统的投入运行保障了气象资料的有效管理,为用户提供了丰富的数据产品服务^[3,5-8]。但是随着气象业务和技术的不断发展,新的需求不断涌现,需要用更加开放的思维来研究气象资料服务系统,丰富和深化服务形式。本文在分析现有服务系统的特点和专业用户需要的基础上,详细介绍一个基于元数据技术的、针对专业和气象业务用户的、能快速下载数据的新型气象资料服务系统的设计与实现。该系统实现了对国家级基础气象资料库中的数据文件以目录树的方式提供用户浏览,并可基于通用下载工具实现快速下载(对于以二维表形式存储的数据服务可以通过传统的数据服务系统进行服务),有效满足了专业用户和业务用户对数据服务的需求。

1 气象资料服务系统特点

气象资料服务系统一般由 3 个系统组成:数据

库系统、应用系统和认证系统^[5]。数据库系统承担数据的存储管理,存储方式一般有两种数据:数据库二维表的存储方式和文件形式。气象信息的多样性、格式的复杂性以及应用的时效性,决定了数据库系统中数据存储方式选择的复杂性。要确定信息的存储结构,就需要将存储方式与资料的格式、检索服务形式相结合,统筹考虑。例如:对于观测资料,用户一般需要提供某个要素的检索,对这类信息采用数据库表方式管理更合适;而对于一些二进制格式的数据或图像及大文本文件,用户一般需要某个完整的资料文件或整个图像,对这类资料采用文件方式管理比较合适。

应用系统主要实现两类服务:数据发现(data discovery)和数据访问(data access)^[1,9]。例如在气象科学数据共享服务平台中,提供的数据导航功能包括数据分类目录导航、关键词搜索导航、相关数据集信息导航和“一站式”搜索等。数据访问即数据检索下载,也提供多种服务,例如数据记录查询、文件下载和数据定制等。一个检索服务首先要选择查询条件,然后完成数据的本地下载。在检索页面上往往提供多个可选择的条件(如图 1 所示)。

近几年,随着 Web 技术的发展,基于 Web 的气象资料服务系统在国家级和省级气象业务部门逐步



图 1 一个基于 Web 的数据检索页面

Fig. 1 One Web page of meteorological data retrieval

建立起来,例如气象科学数据共享服务网和国家级气象资料存储检索系统,前者成为了气象科学数据共享服务的主窗口,后者也逐步成为国家级业务部门的主要数据支撑系统^[4]。目前运行的基于 Web 的气象资料服务系统基本上能够提供多入口的导航和灵活的查询等功能,这些功能对于非专业用户具有较强的实用性,用户可以准确定位自己所需的数据资源,在定位数据资源后,可依据系统提供的多条件查询系统浏览和下载数据文件,同时获取丰富的说明信息。但对于专业用户和气象业务用户,以及频繁使用气象数据文件的用户来说,这些气象资料服务系统显得过于繁琐,同时效率也不高。气象资料服务系统应能够为专业用户和气象业务用户提供如下功能:①简单和清晰的数据文件目录导航功能;②快速下载功能,同时支持对数据文件的批量下载;③安全管理功能(能够示踪用户行为,进行统计分析)。即这个服务系统既要具有使用的方便性和查询的简洁性,同时又能实现数据文件的快速和批量下载。针对该需求设计和开发了一个新型气象资料服务系统——气象数据文件快速下载服务系统。

2 系统设计与实现

2.1 基于 MVC 设计模型的体系结构

气象数据文件快速下载服务系统采用 J2EE 架构和 MVC(model views control)设计模式^[10-13]。

MVC 设计中,模型(model)层只关注数据的存储和存放,表现(view)层则关心对业务逻辑的界面展示,而控制(control)层则是一个协调的主动者,对模型和表现这两个被动角色进行调度,从而实现业务的流转。整个系统按照层次的体系结构分为:表现层、控制层以及业务逻辑层和数据库层,各层相互配合,完成系统功能。每层由一组相关的类或组件构成,共同完成特定的功能。层与层之间存在自上而下的依赖关系,即上层组件会访问下层组件的应用程序接口(API),而下层组件不应该依赖上层组件;每个层对上层公开 API,但具体的实现细节对外透明。当某一层的实现发生变化,只要它的 API 不变,不会影响其他层的实现。该系统表现层提供数据集的树形结构和数据集内容的展现;控制层为业务流程控制层,实现将用户输入信息翻译为系统信息,并根据用户输入或指定的内容选择合适的数据文件;业务逻辑层处理业务逻辑,完成统一存储管理;数据库层记录并保存用户操作轨迹,提供数据信息。

2.2 系统功能设计

气象数据文件快速下载服务系统由前下载子系统 and 后台管理子系统两部分组成。前下载子系统的设计充分考虑了业务用户使用方便性和简洁性,后台管理系统要实现业务管理的便捷性和系统的可扩展性。由于本系统支持大批量数据的快速下载,在为内部用户提供便捷的大批量数据获取方式的同时充分考虑了数据的安全性。

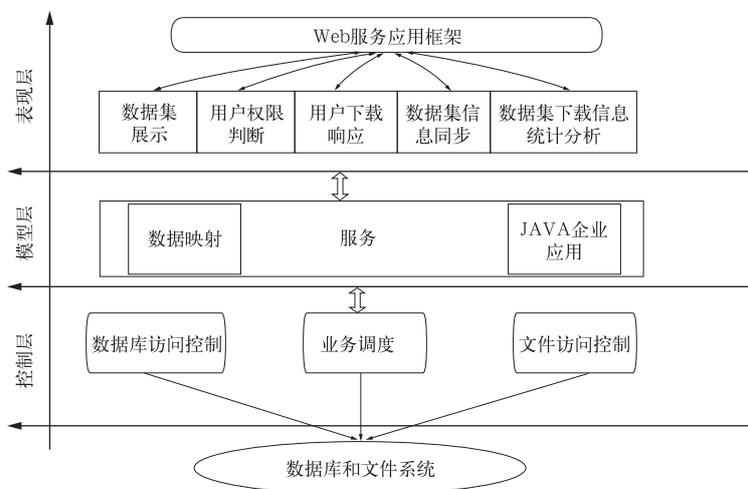


图 2 系统结构示意图

Fig. 2 Framework of Meteorological Data Files Quick Download Service System

考虑到用户的使用习惯,前台下载子系统的数据目录浏览页面以 MS Windows 资源管理器为参照进行页面设计,为了方便用户能一次查询看到多个数据集或者某个数据集下多个子目录中的文件,该系统在 MS Windows 资源管理器功能之上增添了新的功能:用户可以同时选择多个目录,然后一并列出各目录下

的所有文件(图 3)。在浏览或选择每个数据集的时候,系统提供该数据集的相关文档。考虑到专业用户和气象业务用户会频繁地调用大批量数据文件,该系统不仅提供单个文件下载,同时也提供利用下载工具支持批量下载的功能,并且支持断点续传,这将使用户获取数据的过程变得非常便捷高效。



图 3 数据文件目录浏览页面

Fig. 3 Web page of data file catalogues browse

为了实现系统管理的便捷和功能的可扩展性,后台管理系统的设计和开发采用了元数据技术,实现功能模块和栏目的自由添加。后台管理系统实现的功能有:用户管理、角色管理、菜单管理、配置管理、行为查询、同步管理、信息统计等。其中用户管理和角色管理实现用户和角色的增、删、改,以及权限的设置功能,菜单管理功能使得后台管理功能实现可配置和可扩展,基于该功能可以实现后台管理功能展现方式的动态调整。此外当系统需要增加新的管理功能时,只需在此管理功能的页面上配置新功能的名称和接入点等信息,则新功能就能自动部署到后台管理系统中。行为查询和信息统计实现对用户访问行为的查询、统计及输出功能。

3 关键技术研究

3.1 采用 MD5 算法保障数据的安全性

由于底层数据遵循了一定的存储规范,数据存储具有一定的规律性,为了保证数据的安全,防止非法用户通过猜测非法获取数据文件的下载地址,因此需要对下载地址进行加密处理。

在现阶段存在着两种加密方式,单向加密和双向加密。双向加密是加密算法中最常用的,它将可以直接理解的明文数据加密为不可直接理解的密文数据,在需要时可以使用一定的算法将这些加密以后的密文解密为原来可以理解的明文。单向加密刚好相反,只能对数据进行加密,没有办法对加密以后的数据进行解密。为了保证数据的安全性,系统采用单向加密算法 MD5 对信息进行加密。

MD5 (Message-Digest Algorithm 5)是在 20 世纪 90 年代初由 MIT 的计算科学实验室和 RSA Data Security Inc 开发的,经 MD2, MD3 和 MD4 发展而来。Message-Digest 泛指字节串(message)的 Hash 变换,就是将一个任意长度的字节串转换成一定长的整数(128 bit),成为一个不可逆的字符串,这样即使用户看到源程序和算法描述,也无法将一个 MD5 值变换回原始的字符串,从而保证数据的安全^[14]。

在气象数据文件快速下载服务系统中,MD5 技术除了用在数据下载 URL 的加密上,用户密码的加密也使用了 MD5 算法,有效地保证了用户密码的私密性。

3.2 通过 RANGE 参数处理支持多线程断点续传

因为该系统要支持大批量的文件下载,为了避免因为网络、系统环境等外部因素造成下载异常中断,系统需要支持断点续传。该系统是通过解析 HTTP 协议中的 RANGE 参数实现断点续传,同时基于此功能实现了多线程并行下载^[15]。

在 HTTP 协议中,请求方会通过 RANGE 参数告诉服务器数据下载的起始位置,同时发出断点续传的请求,如下:

```
GET /down.zip HTTP/1.0
User-Agent: NetFox
RANGE: bytes=2000070-
Accept: text/html, image/gif, image/jpeg,
* ; q=.2, */*; q=.2
服务器在收到这个请求以后,返回信息,如下:
206
Content-Length=106786028
Content-Range = bytes 2000070-106786027/
106786028
Date=Mon, 30 Apr 2001 12:55:20 GMT
ETag=W/"02ca57e173c11:95b"
Content-Type=application/octet-stream
Server=Microsoft-IIS/5.0
Last-Modified=Mon, 30 Apr 2001 12:55:20
GMT
```

通过这样的消息传递,可保障中止的文件下载继续进行。

3.3 增量同步

为了保证用户通过本系统访问的数据为最新,需要定时在气象数据文件快速下载服务系统与业务数据库之间进行数据同步。系统采用增量同步方式提高同步效率,它是通过比较各类数据的最大 ID 号来实现数据文件的增量同步。当数据文件进行了更新或者修改,将触发同步调度存储过程,自动完成该数据的同步,并同时生成数据树形结构^[16-17]。系统利用了 ORACLE 系统提供的 Job 定时调用同步调度存储过程技术。

使用 Job 调度具有很好的效果,它允许用户提前制作增量同步任务单,使其能在指定的时间点或时间段内自动执行一次或多次。由于任务在数据库中被执行,所以执行效率很高。Job 还提供了灵活的处理方式,可以通过配置,安排任务在系统用户访问量少的时段内执行,能极大地提高运行效率。

3.4 目录导航

根据内容属性,气象数据分为 14 大类,每大类包括许多数据集产品,这些数据集产品分别存储在不同的数据目录下,为了给用户在气象数据文件快速下载服务系统中实现文件的快速按目录浏览,提供目录树方式的数据导航,需要事先分析每个文件的存储结构并存储在数据库中。在系统设计中,首先建立数据文件索引表(FILES 表),FILES 表的 FDIR 字段存储了文件存储的路径。在 FILES 表中存储管理的记录树将会达到百万级并不断增长,为了方便、快速展现数据的层级从属关系,设计一个树形结构关系表来展示数据的从属关系。该表通过数据集 ID、结点 ID、父结点 ID、本层排序标识、叶子结点、显示等字段来描述数据的从属关系,并以树形结构展示。这种设计结构可以从任一个结点找到该结点的父结点和该结点全部子结点,而且只有当该结点是叶子结点时再转向文件存储表获取文件。通过这种结构设计提高了系统的展示速度并减轻了数据库的压力。

3.5 安全配置

由于气象数据资料的安全性要求,系统需采用必要的安全方式对用户访问、下载和管理进行授权检查,拒绝未经授权的各种访问方式。而仅仅通过用户名、密码的方式不能保证授权的安全,系统采用用户名、密码和 IP 地址授权范围的混合方式来进行授权检查以保证数据的安全性。由于整个系统采用 B/S 架构,HTTP 协议本身已经包含了用户的 IP 地址信息,从而在服务器端非常容易就可以通过 HttpServletRequest 类中的 getRemoteAddr 方法方便地获取到用户 IP 地址信息,如果判断该地址在访问的许可范围内,表明访问用户已经获得了系统授权,便可以根据登录权限来为用户提供相应的服务,通过这种方式无需用户经过其他特殊的软件或配置就可以为系统增加更多的信息安全保护。

4 结论与展望

气象数据文件快速下载服务系统是对气象资料服务方式的一种全新设计,也是对目前气象系统两个大型数据管理系统国家级气象资料存储检索系统(为行业内服务为主)和气象科学数据共享平台(对外服务为主)的有益补充,与这两个系统相比,气象数据文件快速下载服务系统有如下特点:

1) 用户可以利用下载工具,快速大批量地下载气象数据文件。

2) 基于系统提供的文件目录导航功能,用户可以按照数据分类和时间特征逐级访问数据,并灵活地进行选择和检索。

3) 实现了与气象数据收集系统的同步更新,用户可以访问下载最新的气象数据产品。

系统发布运行后,许多业务用户通过其下载了大量气象资料,部分用户还利用系统提供的数据访问接口与其数据处理系统进行对接,开展数据分析和数据产品研制工作,有力地推动了预报、预测和灾害监测业务的发展。

参考文献

- [1] 王国复,李集明,邓莉,等. 中国气象科学数据共享服务网总体设计. 应用气象学报, 2004, 15(增刊): 10-16.
- [2] 李集明. 基于 Internet 面向社会公益性共享的气象科学数据存储检索策略研究. 气象科技, 2007, 35(4): 590-592.
- [3] 李集明,沈文海,王国复. 气象信息共享平台及其关键技术研究. 应用气象学报, 2006, 17(5): 621-628.
- [4] 李集明,王国复. 气象数据库系统总体设计综述. 气象科技, 2007, 35(增刊): 1-5.
- [5] 王国复,孙超,何小明,等. 气象目录服务系统的设计与实现. 地球信息科学学报, 2009, 11(1): 24-29.
- [6] 沈文海,赵芳,高华云,等. 国家级气象资料存储检索系统的建立. 应用气象学报, 2004, 15(6): 727-736.
- [7] 陈学君,王国复,陶建红. 基于 Web 西北地区干旱气象资料共享服务分系统. 应用气象学报, 2004, 15(增刊): 161-164.
- [8] 邓莉,王国复,孙超. 基本气象资料共享系统建设. 应用气象学报, 2004, 15(增刊): 33-38.
- [9] 王国复,涂勇,王卷乐,等. 科学数据共享中的元数据技术研究. 中国科技资源导刊, 2008, 40(1): 30-36.
- [10] Buschmann F, Mcunier R, Rohnert H, et al. Pattern-oriented Software Architecture: A System of Patterns. John Wiley & Sons Ltd, 1996.
- [11] 陆荣幸,郁洲. J2EE 平台上 MVC 设计模式的研究与实现. 计算机应用研究, 2003, 20(3): 144-146.
- [12] 费翔林. 基于 MVC 模式的应用软件开发框架研究. 计算机工程与应用, 2005(30): 102-111.
- [13] 李斌,陈钟荣. J2EE 平台技术在气象信息系统中的应用研究. 南京气象学院学报, 2006, 29(2): 274-281.
- [14] 王金柱,李元诚. MD5 算法在 J2EE 平台下用户管理系统中的应用. 计算机工程与设计, 2008, 29(18): 4728-4764.
- [15] 程晓伟,田东风. FTP 超大文件断点上传技术研究. 电脑知识与技术, 2009(6): 1344-1345.
- [16] 盖九宇,张忠能,肖鹤. 分布式数据库数据复制技术的分析与应用. 计算机应用与软件, 2005, 22(7): 36-38; 41.
- [17] 曾剑鹿,李宏. 基于消息队列和 XML 的数据同步技术研究. 电脑知识与技术, 2009(10): 2600-2602.

Meteorological Data Files Quick Download Service System

Gao Feng¹⁾ Wang Guofu¹⁾ Yu Wen¹⁾²⁾ Feng Mingnong¹⁾ Sun Chao¹⁾ Luo Qi³⁾

¹⁾ (National Meteorological Information Center, Beijing 100081)

²⁾ (Beihang University, Beijing 100191)

³⁾ (Gansu Provincial Meteorological Information Center, Lanzhou 730020)

Abstract

The development of meteorological data service system is summarized as 4 stages: Stand-alone file retrieval stage, the interface calling stage, database system stage and distributed data service stage. The current service system can meet the needs of most users querying data, but for meteorological professional and operational users, the system has certain limitations. These users call for a new data service system that offers fast bulk download meteorological data services, which should also be established to make up for the deficiencies of existing meteorological data-sharing service system.

According to operational requirements, a new meteorological data service system is designed based on metadata technology, through which professional and operational users can quickly download batch of real-time meteorological data. The system is based on front-end application systems and background management systems to achieve a bulk download, file directory navigation, data synchronization update, multi-

