

气象数据存贮管理系统^①

窆以文, 卢 俐, 刘旭林, 张代萍

(北京市气象信息中心, 北京 100089)

摘 要: 为解决当前北京市气象局历史气象数据庞大, 存储结构欠合理, 存储格式不统一, 数据管理不明晰, 以及科研人员、气象预报人员获取历史气象数据难两大突出问题, 提出了气象数据存贮策略, 并以 B/S 结构框架, 开发了气象数据存贮管理系统。该系统为业务和科研人员在数据获取上提供了一个便捷通道, 同时, 又将数据统一、有序管理起来, 实现了气象数据的存储、管理、以及服务的整合和一体化。本系统的实现将为气象数据存贮管理提供了一个很好地样例。

关键词: 结构化查询语言; 气象数据; 海量存贮系统

Meteorological Data Storage and Management System

DOU Yi-Wen, LU Li, LIU Xu-Lin, ZHANG Dai-Ping

(Beijing Meteorological Information Center, Beijing 100089, China)

Abstract: At present, the management and service of non real-time meteorological data at Beijing Meteorological Bureau have two problems. One is that storage structure of meteorological data is not reasonable, the format of storage files is not uniform, and data management is not clear. The second is it's difficult for researchers and weathermen to obtain the needed meteorological data. To solve the above problems, a Meteorological Data storage and management System has been developed with ASP language. The system structure is based on Microsoft Server 2003+Microsoft SQL Server 2005+Web. This system open a convenient channel for data acquisition. On the other hand, it integrates all kinds of meteorological data orderly. In conclusion, the meteorological data storage and management System makes the meteorological data integration on storage, management and service come true. This system is a good example of the management of the meteorological data.

Key words: SQL; meteorological data; mass storage system

SQL Server 是微软开发的一个数据库管理系统, 功能非常强大。它具有良好的可收缩性、安全性、易用性, 以及大容量的管理能力(支持 TB 级数据量), 广泛应用于大中型数据库系统中^[1-3]。SQL Server 2005 版本不仅是迄今为止 SQL Server 的最大发行版本, 而且是最为可靠安全的版本。

随着气象观测技术的发展以及观测硬件设备的更新, 地面、高空气象观测站点和观测要素均在逐年增加, 每天新生成的观测数据以千兆级数量单位在增长。

同时, 气象卫星每天也在获取各种各样的遥感观测数据。另外, 由于气象预报业务的需要, 每天也有来自全球各地的数值预报产品经由通信传输系统接收。此外, 由于工业的发展、城市的扩张, 对大气成分的监测日益受到重视, 获取了多种相关观测资料。因此, 气象数据具有量大、种类繁多的特点。

当前, 各种实时气象数据和产品, 基本满足了气象预报和科研人员的需要。但是, 海量的历史气象数据存储结构欠合理, 存储格式不统一, 部分气象数据

① 基金项目:北京市气象局基金(BMBYWFZ2010001)

收稿时间:2010-11-01;收到修改稿时间:2010-11-24

还存在多次保存的冗余现象, 缺少数据到达和使用情况监控等。如何将这些珍贵的气象数据整合到一起, 分门别类妥善保存, 一方面方便气象数据管理人员的管理以及数据服务, 另一方面方便科研人员以及气象预报人员最快最便捷获取所需的各种历史气象数据, 让气象数据产生最大的社会服务价值, 是一个亟待解决的重要问题。

为解决以上两大突出问题, 基于 SQL Server 2005 数据库系统, 北京市气象局开发了气象数据存储管理系统, 实现了气象数据的存储、管理, 以及服务的整合和一体化。

1 数据管理策略

1.1 数据存储策略

数据的存贮主要采用数据库表和文件两种形式即元数据和结构化的数据采用数据库形式, 非结构化数据采用文件方式存储。在综合考虑现有设备的存储能力和调取数据的效率基础上, 将所有数据按用户使用的频率和单个文件的大小将其分为 4 个等级, 数字小的等级最高。用户常用的数据定为 1 级, 不常用的或文件字节数较少的定为 2 级, 常用但文件字节数超大的定为 3 级, 根本没人用的定为 4 级。

每种数据采用一天一个压缩包的保存方式 (rar 文件), 如果一天仅有一个文件, 则存储为原始文件。数据在盘阵上存储的目录结构定为 4 级, 第一级目录为资料类型, 第二级目录为资料子类型, 第三级目录为数据生成的年份, 第四级目录为数据生成的月份。这样的目录结构清晰明了, 便于日后对数据的管理和维护。

综合考虑数据下载速度与存储能力, 将 1 级数据长期保存在盘阵上, 2、3 级数据只在盘阵上保留 5 年, 4 级数据只在盘阵上保留 1 年。超过在盘阵上存储年限的数据一律下线, 即将数据刻录入光盘或存入磁带。

另外, 将有天气过程日期的数据和重要事件日期节点的数据长期保存在盘阵上, 为气象科技人员分析总结天气过程提供方便。

1.2 数据分发策略

在资料的分发上综合考虑资料时效和传输效率。集中资料的存储, 减少资料存放处, 保持数据的一致性, 方便用户获取资料。目前的数据分发模式有三种, 一是将数据放公共数据区, 由用户自行调取; 二是由

信息中心直接向业务单位的应用系统推送数据; 三是特殊资料服务由人工处理。对不同级别的用户赋予不同的下载/接受资料的权限。

1.3 数据分发策略

建立资料整编审核制度, 对实时资料和历史资料进行审核, 包含设备维护及故障情况等电子文档; 制定资料存贮规章制度, 所有的探测资料都要按一定形式实时归档; 定期制作常规数据集。

1.4 数据安全策略

提供历史数据服务需要对文档加密, U 盘贴 RFID 标签, 设定数据使用范围, 有效期等; 实时气象资料服务需要建立信任体系如 CA 认证; 除了管理好日常的常规备份, 还要做好异地数据灾难备份。

2 系统结构及功能

本系统采用 B/S 构架, 以 Microsoft Server 2003+Microsoft SQL Server 2005+Web 为系统框架, 总体由后台数据分拣程序、气象数据查询下载网页、后台 FTP 实时监测管理程序、系统后台管理四部分组成, 图 1 是该系统总体框架流程图:

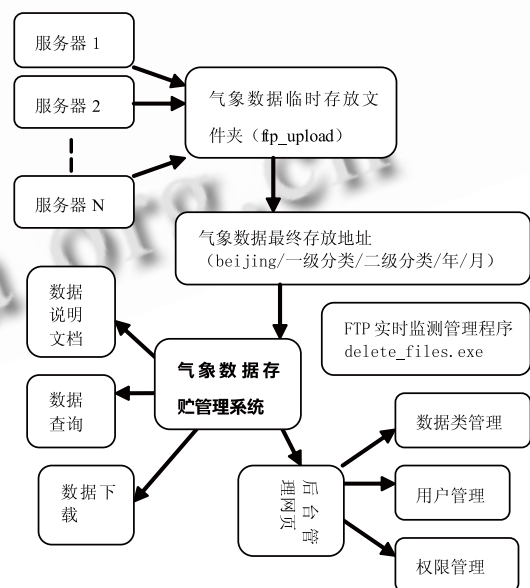


图 1 “气象数据存储管理系统”总体框架流程图

2.1 后台数据分拣程序

每天从各种渠道获取的大量气象数据, 以一天一个文件形式, 按照约定的文件名 (表 1, 文件名包含文件类型及日期两种信息), 通过 FTP, 从各个服务器传到存放数据的服务器硬盘, 临时放入一个指定的文

文件夹 (ftp_upload), 等待按照类别分拣入最后存放地址。数据分拣程序 (sorting_data.exe) 用 VB 语言编写, 该程序首先执行一个 bat 小程序, 将 ftp_upload 文件夹中的所有文件的文件名写入 txt 临时文本, 然后逐一检查每个文件名, 只将当前日期 2 天前的所有文件转移到每类数据存储的最终文件夹, 按照 “beijing/资料类型/资料子类型/年/月” 的路径顺序存放。近 2 天的数据文件则继续保留在 ftp_upload 文件夹, 等待下次入库。数据入库有一个时间上的滞后, 主要是因为部分当天生成的数据文件会在当天晚些时候或者第二天更新该文件, 为了保证最终存储的数据是最完整最新的, 做了一个时间上的调整。每转移一个合格文件, 则将该文件的文件名和存放地址写入 SQL Server 2005 数据库对应中, 为后续数据管理及查询服务。在这个数据库里, 以每类数据的代码为表名 (表 1) 建表。

若 ftp_upload 文件夹中有不在计划入库范围内的文件 (即表 1 中找不到对应的文件名), 则在 sorting_data.exe 执行中自动删除。当所有数据分拣完成后, 转移的每类数据的文件数追加写入到 log 文本, 方便管理员检查与核对。

表 1 数据分类

资料类型	资料子类型
地面气象资料	北京市自动站小时观测资料
	北京市自动站道面观测资料
	北京市自动站小时降水观测资料
	北京市闪电定位观测资料
	地面天气报
	ADTD 雷电定位系统数据
	北京市自动站分钟观测资料
	北京市自动站 5 分钟观测资料
	北京市自动道面监测分钟观测资料
	北京市自动道面监测 5 分钟观测资料
高空气象资料	交管局道面资料
	路政局道面资料
	大监网自动站 Z 文件
	区域自动站 Z 文件
	水利局降水数据
	北京市 GPS 观测资料
	北京市观象台风廓线观测资料
	北京市延庆风廓线观测资料
	北京市海淀风廓线观测资料
	北京市上甸子风廓线观测资料
风廓线观测资料(1h)	
风廓线观测资料(30min)	
风廓线雷达基数据(原始数据)	
风廓线雷达基数据(傅里叶变换)	
风廓线温度数据	
常规高空报报文	

农业气象资料	航空报报文
	观象台微波辐射计时序图
	车道沟微波辐射计时序图
	观象台微波辐射计廓线图
	车道沟微波辐射计廓线图
	土壤墒情
	北京市紫外线辐射资料
	日照
	观象台辐射数据
	北京市 LAPS 数据文件
气象辐射资料	欧洲中心数值预报文件
	日本传真图
	T213 数据
	Grapes 产品
	T639 产品
	日本数值格点报
	自动站加密数据
	北京市观象台大气成分观测资料
	北京市上甸子大气成分观测资料
	负离子
数值分析预报产品	花粉
	酸雨
	风沙流
	沙尘暴
	宝莲污染物观测数据
	昌平污染物观测数据
	朝阳污染物观测数据
	华北区域沙尘暴数据
	上甸子太阳光度计数据
	上甸子温室气体数据
大气成分相关资料	北京市观象台气象灾害资料
	雷达基数据
	气象灾害资料
	雷达 Pub 产品
	气象雷达资料
	雷达图片
	雷达状态文件
	静止卫星风云 2C (图片)
	静止卫星风云 2C (gpf 格式)
	静止卫星风云 2D (图片)
静止卫星风云 2D (gpf 格式)	
气象卫星资料	Micaps 格式静止卫星低分辨率图片
	Micaps 格式静止卫星产品 (2C,E)
	Micaps 格式静止卫星产品 (2D)
	2E 星原始数据
	2D 星原始数据
	Micaps 地面高空报
	典型天气资料
	其他资料
	区域自动站 A 文件
	区域自动站 J 文件

注:

(1) 典型天气资料, 主要为预报员服务。它集中提供出现典型天气的某日 (强对流、大雨、中雨、雪、大风、雾霾、寒潮等) 的自动站小时观测、雷达图、

卫星云图、风廓线以及 Micaps 格式的地面高空、欧洲数值预报数据;

(2) 2009 年 11 月 30 日开始, FY2E 星取代 FY2C 星, 对于图像产品文件, FY2C_images_YYYYMM DD.rar, C 代表的实际为 E 星的图像产品; 对于 gpf 产品文件, 文件名变更为 FY2E_gpf_YYYYMM DD.rar。

(3) 每一种资料都有固定的编码和标准的文件命名。

2.2 气象数据存贮管理系统网页

“气象数据存贮管理系统”网页用 ASP 语言编写, 以 Microsoft SQL Server 2005 为后台数据库, 网页以 ADO 的方式访问数据库^[4]。在网站的首页, 展示了当前可供查询下载的所有数据类别, 以及所存的该类数据的起始和结束时间。这些信息均从数据库中提取, 随数据库的更新而更新。

可供用户网上下载的数据, 则是根据资料等级, 从表 1 中选出。不能网上获取的资料, 则为用户提供离线服务。另外, 根据用户的工作性质, 也赋予用户相应的等级(A、B、C、D 四级, A 级权限最大), 不同等级用户可下载数据的种类是不同的。

该存贮管理系统主要有两大特色。一是所有的数据类型均有非常详细的说明文档链接, 可实时查看或下载, 方便用户预先了解所下载数据文件的具体内容, 并在数据下载后可立即使用。改变了以往数据下载系统只能下载数据, 却因为缺乏相关的数据说明文件而使用户就算拿到了数据, 也难以使用或者无法使用的局面。二是, 为提高数据下载速度, 采用动态 ftp 的形式为用户提供所需数据。当用户给出查询条件后, 系统检索 SQL Server 2005 数据库, 将符合条件的文件拷贝到一个可用的动态的 FTP 文件夹。检索拷贝完成后, 网页告知用户 FTP 用户名和密码, 用户可在一定时间段内(48 小时)下载。48 小时后, FTP 的密码被更改, 该文件夹被清空。本系统还设定, 同一用户在 24 小时内的各种查询文件, 均放入同一文件夹, 这样大大有效利用了每一个 ftp 目录, 同时也方便用户下载所需的各种数据。

2.3 后台 FTP 实时监测管理程序

前一节提到的动态 FTP 的实现, 需要一个对用户使用的 FTP 的监测管理程序。在存储数据的服务器上安装 FTP 服务器软件 Serv-U, 该软件具有功能

完备、安全性良好、界面简单友好、易于使用的特点, 因此使用广泛^[5]。利用 Serv-U 的 ODBC 功能, 用户的 FTP 信息存储于数据库, 这样 FTP 用户密码可方便更改。

在本系统中, Serv-U 6.2.0.0 和 SQL Server 2005 Server 2005 相结合, 在 Serv-U 中创建一个新的域(ftp.bjdata.com), 并将域类型选择为“store in ODBC database”。在该域中设定 100 个用户(当前暂定), 指定到 100 个互相独立的主目录, 并锁定主目录。这 100 个用户的 FTP 信息存储于 SQL Server 2005 数据库的表“ftp_users”。

在数据库里, 另建两个表, “ftpuser”和“timer”。“ftpuser”存储每个 FTP 主目录的使用状态。“timer”记录正在使用的 FTP 主目录信息, 包含“文件夹, 地址, 开启时间”三个字段。当用户查询下载时, 系统根据表“ftpuser”中的主目录使用信息, 将数据拷贝到一个未使用的主目录, 该主目录的使用状态在表“ftpuser”中被更改。同时, 在 timer 表中生成一条有关该主目录开始使用的记录。后台 FTP 实时监测管理程序(delete_files.exe)实际监测的即为 timer 表。该程序用 VB 语言编写, 借助于 VB 中的 timer 控件, 监测程序每 30min 检查一次 timer 表, 如果开始使用时间与当前时间比较, 超过 48 小时, 则清空对应主目录中的所有文件, 更新表“ftpuser”中该主目录对应 FTP 用户的密码(MD5 算法实时生成一个新密码)及使用状态, 并在 timer 表中删除该主目录的使用记录。由此该主目录完成一次使用, 并等待进入下一次使用。

FTP 实时监测管理程序一直运行于服务器, 这是一个小程序, 长期运行并不影响服务器速度。使用动态 FTP, 有效利用了服务器硬盘空间, 不过多占用资源, 并且实现 FTP 的循环可再用。

2.4 气象数据存贮管理系统后台管理网页

为了方便对下载数据的用户进行管理, 以及实时调整每类数据的下载权限, 为数据管理和服务人员开发了“气象数据存贮管理系统后台管理”页面。网页用 ASP 语言编写, Microsoft SQL server 2005 作为后台数据库。管理分三部分: 数据类管理、用户管理、权限管理。数据类管理, 包括增加新的数据类, 每类数据在 ABCD 四级用户中的下载权限; 用户管理, 主要是查看用户下载数据的详细信息, 作为修正数据存储策略的依据, 并且必要时冻结用户账户。权限管理,

则主要是根据每个用户的工作性质,分别赋予不同的用户等级(A级最大,可下载数据种类最多,D级最小,也是用户注册时查询下载系统自动赋予的最初权限等级);

3 结论

“气象数据存储管理系统”为我局业务和科研人员在数据获取上提供一个便捷通道,大大减轻了数据管理人员在这方面的重复手工操作,提高了工作效率。同时,又将数据统一、有序管理起来,实现了气象数据的存储、管理、以及服务的整合和一体化。

需要指出的是,存储数据的服务器、“气象数据存储管理系统”、SQL Server 2005 以及 Serv-U 均部署于同一服务器,可实现最快的数据查询下载服务,同时也方便管理。该系统目前处于初步运转阶段,随着数

据量与种类的与日增加,服务器性能的承载能力如何,还需要进一步的测试并以此调整系统。

参考文献

- 1 Nielsen P.赵子鹏,袁国忠,乔健译.SQL Server 2005 宝典.北京:人民邮电出版社,2008.
- 2 Sack J.朱晔,金迎春译.SQL Server 2005 范例代码查询辞典.北京:人民邮电出版社,2008.
- 3 张淑萍.地级市气象数据库的设计与实现.农业网络信息,2009,8:73-75.
- 4 王海龙,杨忠霞,陈朝辉,吕东柏,尹英汉.地级市气象信息网络数据库系统建设及应用.气象环境科学,2009,32(增刊):311-314.
- 5 王婉,姚博,周鸣.FTP 服务器的建立及客户端的使用.河南气象,2006,1:75-76.

(上接第 187 页)

6 结语

本文在介绍无线传感器网络相关知识的基础上,以 CC2530 模块为硬件基础,外加 CC2591 功放模块,基于 ZigBee2006 通讯协议进行组网。文中阐述了组成无线传感器网络所需的传感器节点,协调器节点,PC 管理机等的软硬件设计方法及流程图等。经测试采用 TI 公司第二代片上系统 CC2530 设计的节点间的通讯距离远大于第一代产品,通讯更可靠。

参考文献

- 1 ZigBee Alliance. ZigBee Specification 2006. [2010-08]. <http://www.zigbee.org>
- 2 李文仲,段朝玉.ZigBee2006 无线网络与无线定位实战.北京:北京航空航天大学出版社,2008.23-30.

- 3 CC25912.4GHz RF Front End (Rev.A) [2010-08]. <http://focus.ti.com/docs/prod/folders/print/cc2591.html>
- 4 CC253X User's Guide [Rev.A].[2010-06-25].<http://focus.ti.com/docs/prod/folders/print/cc2530.html>
- 5 IEEE 802.15.4-2003Std: Wireless Medium Access Control (MAC) and physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Network (LR-WPANs). [2010-08]. <http://www.ieee802.org/>
- 6 Suggestions for the improvement of the IEEE802.15.4 standard. [2010-08]. <http://www.ieee.org>.
- 7 台宏达,田国会,宋保业,等.ZigBee2006 协议栈的无线传感执行网络构建.单片机与嵌入式系统,2009,7:35-38.