

大城市气象服务信息系统数据存储 体系框架设计与实现

王京丽 谭晓光 张德政

(北京城市气象工程技术研究中心,北京 100089)

李 伟

(中国气象局大气试验基地,北京 100089)

摘要 海量气象数据的存储、管理以及信息服务的实现一直是气象信息处理中的关键问题。文章系统介绍了我国大城市气象服务信息系统技术框架中气象数据种类和数据格式、数据结构和数据存储体系以及各主要数据类型的管理和存取方式。

关键词 气象数据管理体系 数据格式 数据存储 MICAPS

引言

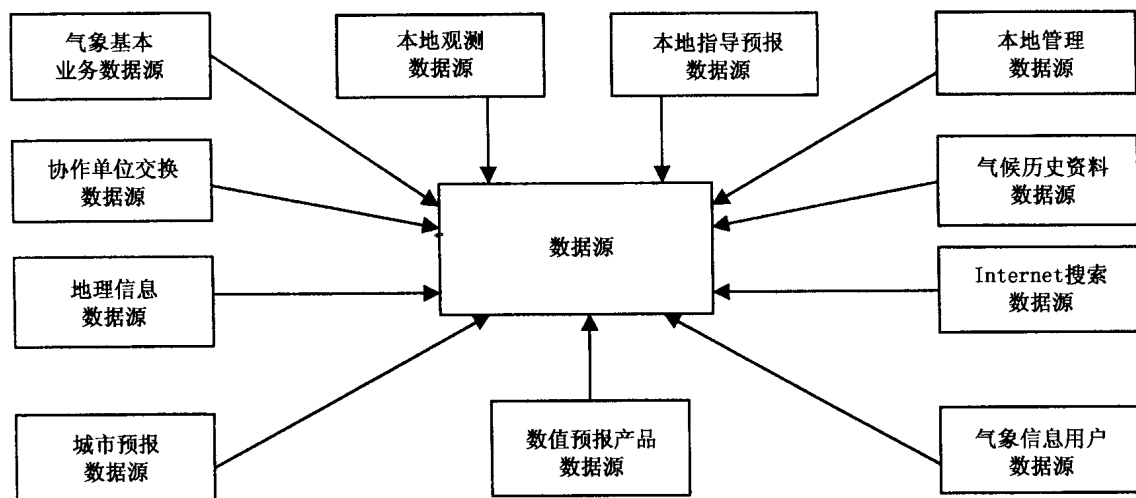
气象数据是气象预报和进行气象科学研究的基础,因此,实时数据、历史数据以及相关的气象数据存储与管理自然就成为各类气象应用系统的重要组成部分。同样在大城市气象服务信息系统中,数据存储与管理框架的研究是系统建设的重要内容。同时,气象数据作为一类科学数据,专业化、规范化的数据存储模式是高可靠、大容量的数据存储与管理体制得以高效运行的关键。

大城市气象服务信息系统的主要特点就是充分利用已有大量数据提供各种气象服务。因此对于种类多、范围广、实时性强、数据量大、类型复杂的气象数据,应该选择有效的存储与管理模式。目前常用的 Oracle、Sybase 以及 Informix 等数据库系统,主要适用于商业事务数据的存储管理以及检索查询与分析等。由于气象数据类型的复杂性以及巨大的数据量,使得一般商业数据库的适应性受到很大限制。多方面技术经济分析论证表明,文件系统是较为适宜的存储方式。目前国内外许多科学数据大都采用文件系统作为存储方式,不同的科学数据采用不同

的格式,同时有相应的存储管理体系用以完成数据的读、写以及检索。美国的 A WIPS (Advanced Weather Information Processing System),其格点数据是用 NetCDF 格式存储^[1~2],而云图、雷达数据则是直接存储原始数据文件。我国气象部门统一使用的 MICAPS (Meteorological Information Comprehensive Analyzed Processing System),采用了类似的存储方式。“大城市气象服务信息系统”数据存储与管理体制中考虑采用类似的设计思想,文件格式采用 BIN 格式。数据量大是气象数据的显著特点,以文件形式存储的数据则可以利用先进的文件压缩技术对其进行压缩,从而大大节省了磁盘空间。相应的数据管理系统,可以方便地对数据库进行各项操作,而且只需提供特定的特性字段数据即可方便、快捷地检索到所需数据。

1 数据种类及其来源

大城市气象服务信息系统所涉及到的数据根据来源大约分为十几类,如图 1 所示。这些数据分别来自不同的数据源,每一类数据中又包含多种不同数据。由此可知,气象数据不仅种类繁多,数据类型

图1 大城市气象服务信息系统数据源^[3]

复杂,而且数据量很大,数据之间具有较强的相关关系和时间序列特征。同时,数据的存取与应用逻辑较为单一,数据的管理功能较为简单。因此,对于复杂的海量气象数据采取文件方式进行存储,同时辅以相应的服务软件包对其进行访问和操作是较好的管理方式。

2 数据存储与管理体系

2.1 数据库体系结构

(1)从存储结构角度上,数据库可分为历史数据库与实时数据库,实时数据库自动保存48h实时资料,48h后转入历史数据库。

(2)从数据逻辑角度上,数据是按两层结构进行存储的。一层是索引文件,另一层是数据文件。这就类似于数据库中通常使用的元数据管理方式。物理上数据库与用户工作程序也是相互隔离的,分别安装在不同的机器上,从一定程度上保证了数据库的安全。

上述逻辑上的第一层是元数据,它是各类数据的索引信息表文件,元数据所记录的内容包括数据文件的属性、类别、时间、存放路径、文件名结构、文件存储格式等信息。第二层是数据库中所有的数据文件,约11类,每类数据中又包含不同种类的数据,它们根据种类分别存储在不同目录下相应的数据文件中,所有数据在数据库中均压缩存储。用户通过元数据库中提供的相关信息进行搜索,即可在第二层的数据文件中查询和检索到所需数据。

2.2 各类数据存储结构

上述提到真正的数据文件是根据数据种类分别存

储在相应的目录下的,每类数据文件的数据结构如下:

(1)“9210”实时数据库中的地面观测数据和高空观测数据为每月一个二进制文件。文件结构为开头是2个20字节字符串,分别为要素中文名和英文名。每天的资料以年、月、日、时、总站数开头,后面是各站数据。

(2)台风数据为每年一个二进制文件。文件结构为开头是2个20字节字符串,分别为:“台风路径”和“Typhoon”。后面为各台风的数据。

(3)欧洲、日本、美国数值预报为每月每个要素一个二进制文件。文件开头为2个20字节字符串,分别为:要素中文名和英文名。后面为经度格距、纬度格距、起始经度、终止经度、起始纬度、终止纬度、纬向格点数、经向格点数。后面每天、每层的资料以年、月、日、时、时效、层次开头,然后按先纬向后经向的顺序放每个格点的值。对风要素,先放 u 分量,然后放 v 分量。

(4) T213、HLAFS 高空数据同欧洲数值预报格式。

(5) T213、HLAFS 地面数据同 T213、HLAFS 高空数据格式,只是每天的资料以年、月、日、时、时效开头。

(6)雷达和卫星图像数据是每个时次的数据为一个文件。文件开头为2个20字节字符串,分别为:“雷达数据”和“Radar”或“卫星云图”和“Star”。后面为年、月、日、时次及相关的特性信息和各像素值。

(7)自动站格式为每月一个二进制文件。文件开头为2个20字节字符串,分别为:自动站观测要素中文名和英文名。每天的资料以年、月、日、时、

分、总站数开头,后面是各站数据。

(8) 中尺度模式格点数据格式为每月每个要素一个二进制文件。文件开头为 2 个 20 字节字符串,分别为:要素中文名和英文名、相关的特性信息。每天的资料以年、月、日、时、时效、层次开头,然后按先 x 方向后 y 方向的顺序放每个格点的值。

其它类型的数据格式也是每月一个二进制文件。文件开头为 2 个 20 字节字符串,分别为其中文名和英文名,每天的资料具体格式不再详述。

2.3 数据库管理系统

为了实现数据的安全管理与操作,便于用户访问数据,我们设计了一套专用的数据库服务软件包,用户可通过它实现对数据库的各种访问。如来自不同数据源的数据经服务软件包中的数据处理功能适当处理后以一定的格式自动存入数据库,同时自动生成相应的索引文件,存入元数据文件中;用户若需从数据库中检索数据,或将生成的服务产品写入数据库中,也是通过服务软件包和元数据文件来操作的。其结构如图 2 所示。

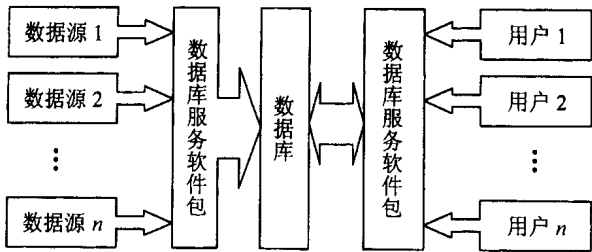


图 2 数据库系统操作流程图

(1) 数据入库方式

数据入库采取如下 3 种方式: ①将各种历史数据经过处理、压缩后存入历史数据库; ②将实时数据定时处理、压缩后存入历史数据库; ③为服务数据自动生成系统及服务产品自动生成系统的数据检索生成必要的中间结果,写入实时数据库。

(2) 数据检索方式

用户对数据库的检索是通过服务软件包来完成的,检索包括数值数据检索、图像数据检索、文本数据检索。

①数值数据检索包括常规地面、常规高空、欧洲数值预报、日本数值预报、美国数值预报、T213 数值预报、HLAFS 数值预报、MM5 数值预报等。用户只需输入检索数据种类、要素名、层次、时间跨度、时

效、特性、返回数组结构等检索参数,即可快速检索出所需数值数据。

用户输入以下参数,即可得到数据检索结果 data 结构:

- sort —— 检索数据种类
- element —— 检索数据要素名
- level —— 检索数据气压层次
- time —— 检索数据时间
- extend —— 检索数据预报时效
- position —— 检索数据区域

②图像数据检索包括雷达、云图、传真图等。用户只需输入检索图像种类、图像源、投影类别、时间跨度、时效、返回结构等参数,即可快速检索出所需图像数据。

③文本数据检索包括预报记录、操作记录、电子会商结果、用户需求、用户反馈意见等。用户只需输入检索文本的种类、文本来源、时间跨度、返回结构等参数,即可快速检索出所需文本数据。

用户输入以下参数,即可得到数据检索结果 data 结构:

- sort —— 检索数据种类
- source —— 检索数据类型名
- time —— 检索数据时间

(3) 数据备份方式

①历史数据库,定期检查历史数据库目录空间,定期备份数据库。

②实时数据库,自动保存 48h 实时资料,48h 后转入历史数据库。

2.4 系统运行环境

本系统数据库服务软件包中的数据处理部分及服务产品中间结果生成部分在数据库服务器上,为 Unix 运行环境。检索部分可在微机上,为 Windows 98 以上系列版本以及 Unix 运行环境。

3 结语

由于包括 Oracle 与 Informix 等著名数据库产品在内的商业数据库,很难适应于气象领域数据的存储与管理。本文根据气象数据类型复杂、数据量大等特点,给出了“大城市气象服务数据库系统”的体系结构、数据结构、数据存取、检索等管理操作的设计与实现。对于类似于气象数据这种单个字段数据量极大的情况使用文件形式建立数据库是最佳方

式。目前该数据库系统已经建成并投入试运行,得到了用户的好评。系统的设计思想、体系结构以及建设经验对于与之相类似的其他学科数据库的建设都具有很好的借鉴。

参考文献

1 Green J L. Advanced Weather Interactive Processing System (AW-

IPS) in a National Center. <http://ams.confex.com/ams/annual2000/techprogram/paper-5928.htm>

2 Bruehl P. NetCDF Software Training Module. <http://stre.comet.ucar.edu/old-sac/srl/data/netcdf>

3 谭晓光. 大城市气象服务信息系统技术报告. 北京:北京市气象局,2001

Design and Application of Data Access Structure of Metropolis Meteorological Service Information System

Wang Jingli Tan Xiaoguang Zhang Dezheng

(Beijing Municipal Meteorological Bureau, Beijing 100089, China)

Li Wei

(Atmospheric Observing Experimental Base, CMA, Beijing 100089, China)

Abstract: The storage and management of a huge quantity of meteorological data and information service are the bottlenecks of meteorological information processing. The details of the data structure, storage frame and data accessing method of the Metropolis Meteorological Information Service System are given. The results are useful for other meteorological application systems and management of similar scientific data.

Key words: meteorological data management, data format, storage and access of meteorological data, MICAPS

《气象科技》开辟“科技快报”新栏目启事

为适应科学技术高速发展的形势,本刊拟开辟新栏目“科技快报”,尽快报道科研、业务、应用等方面的新发现、新思路、新方法、新技术、新经验。该栏目稿件的基本要求为:①内容新颖,文字精炼;②稿件篇幅2000字左右(1页),应包括文题、作者姓名、单位、Email、正文及关键词(暂不需要英文信息)。

本栏目稿件系详细文摘,不要求详细的科学论证与阐述,突出新颖性、创造性和适用性,允许探讨与争鸣,但不得违背基本科学原理。本栏目稿件采用特殊审编程序,以加快出版速度。以本栏目报道过的思路、方法为基础的后续研究成果可另文在本刊或其他刊物发表。欢迎投稿!

《气象科技》编辑部

2003年10月