

气候资料数据管理技术的研究及应用

花灿华

(国家气象中心,北京,100081)

摘要 文章着重阐述了气候资料数据管理技术的定义和重要性,介绍了气候资料存储内容、气候资料数据压缩方法和各种资料标准记录格式、气候资料存储系统以及数据检索方法。

关键词 气候资料 数据管理 分布式数据库

1 引言

气候资料是国民经济建设、国防建设和科学研究中不可缺少的重要依据,尤其在四化建设中,从工业布局、农业气候区划、气候资源的开发直至国防尖端等部门都需要大量的气候资料进行统计、分析。气象业务和科研更离不开气候资料的使用。

随着时间的推移,科学探测手段和加工手段的发展,气候资料的信息量与日俱增。据估计,近几年我国气候资料每年以90—100GB的速度增长。如何对这样庞大的信息量进行科学组织、管理、存储、使用已成为气象部门各级领导和科技工作者所关注的问题。为了开发气候资料资源的利用,充分发挥气候资料的潜力,以适应四化建设的需要,对此,必须加强气候资料数据管理的业务体系现代化建设,采用先进的数据管理技术。

数据管理技术主要指的是数据的组织、编目、定位、存储、检索和维护等,它是数据处理的核心问题。众所周知,数据管理技术的高低,对数据处理方式和效率有直接影响。同样,气候资料数据管理技术的高低,对气候资料数据处理方式和效率也有直接影响。为了提高气候资料处理的能力和效率,及时准确、

方便可靠地提供服务,必须加强气候资料数据管理技术的研究。多年来,国家气象中心资料室对气候资料数据管理的有关技术进行了认真的研究和应用,取得了一定的成绩,促进了气候资料业务现代化建设的发展,提高了气候资料服务的效益。

本文着重介绍气候资料存储内容、气候资料信息系统规范化、气候资料数据存储系统、气候资料检索方法等。

2 气候资料存储内容的研究

随着近代科学技术的发展、气候研究内容不断扩大和深化,气候资料的概念和内涵已远远超出人们通常所认为的常规观测资料的范围,而应包括与整个气候系统,即大气圈、海洋圈、冰雪圈、陆地和生物圈有关的资料。根据世界天气监视网(WWW)和世界气候资料计划(WCDP)的考虑,经我们的研究,气候资料存储的内容应包括经由全球通讯系统(GTS)传输的直接关系到大气-海洋-冰雪-陆地气候系统的资料,具体有:

(1)高空资料(大气):气压、温度、风向、风速、湿度、水汽含量;

(2)地面气候资料(大气-陆地):降水(雨和雪)、温度、气压、风向风速、云、蒸发、雪(覆

盖、类型、深度、含水量)、水汽含量、日照、辐射、天气现象;

(3)海洋面的水下资料(海洋):海面风、温度、气压、海表温渡、海-气温差、海水温度、海流、蒸发、降水;

(4)冰雪圈资料(海洋和陆地):冰川、大陆冰层、海洋水冰边界、海冰覆盖、厚度、融解与漂浮、雪覆盖和含水量;

(5)辐射收支(大气-陆地):有关的覆盖、类型、高度、厚度或光学厚度、行星辐射收支分量、太阳常数紫外线通量、地面反照率、地表辐射、红外辐射、陆地和冰面温度;

(6)大气成分资料(大气-陆地): CO_2 、 O_3 和其它辐射性活动气体、甲烷、痕量气体、平流层 H_2O 和气溶胶、对流层气溶胶、浑浊度、污染、大气与降水化学;

(7)水文资料(陆-地):地面水(河、湖、水库的水流沉淀物输送/沉积,水的物理和化学性质及温度、冰覆盖的性质和范围)、地下水(水位高度、温度);

(8)土壤与植被资料(陆地):蒸发/蒸散、植物水应力、地表及不同深度的温度、湿度、植物覆盖及变化;

(9)数值预报产品资料。

3 气候资料信息系统规范化的研究

气候资料信息系统规范化是建立数据管理系统的必要准备和基础。为了提高气候资料的统一性、准确性和可利用性,进一步开发气候资源的利用,充分发挥气候资源的潜力,以便实现气候资料数据的互换共享。对此,研究气候资料信息系统的规范化是十分必要的。

3.1 气候资料标准化格式的研究

气候资料标准化格式是气候资料进行交换、存储和计算机进行数据处理的应用格式。随着电子计算机的迅猛发展,越来越多的气候资料将需要并可能在信息化媒体上进行交换。各地气象台站所获取的气候信息不仅自己用,也不仅与临近地区和省交换,而且要进

行全国性和国际性的交换,这就必须对各类气候资料的格式制定标准化格式,以便实现气候资料的互换、共享。

在进行气候资料标准化格式的设计过程中,我们始终遵循如下基本原则:科学性、系统性、可扩性、兼容性、合理性、适用性、紧凑性等;充分考虑了世界气象组织的有关规定,参考了美、英等国的有关文献,并结合我国的具体情况,从需要和可能两个方面进行了深入的对比分析,力求标准格式有广泛的实用性。

目前气候资料涉及的范围极广,不仅包括常规观测的气候资料,还涉及到大量非常规观测气候资料。鉴于目前的条件和资料现状,对常规观测的气候资料标准格式考虑较为全面;对于非常规观测气候资料,由于资料情况比较复杂,仅考虑了部分卫星资料和国内雷达图像资料的标准格式。主要设计了如下几种气候资料的标准格式:

- 1)全国陆地测站地面气象资料格式;
- 2)全球海洋测站地面气象资料格式;
- 3)全球高空气象资料格式;
- 4)全球地面月平均气象资料格式;
- 5)全球高空月平均气象资料格式;
- 6)国内陆地测站地面气象资料格式(数值型);
- 7)国内陆地测站地面气象资料格式(字符型);
- 8)卫星晴空辐射率观测资料格式。

另外,在格点资料方面,世界气象组织建议推广 GRIB 码,即二进制格点资料码。经研究,GRIB 码适合我国使用,可作为我国格点资料方面的标准化格式。在测站方面,世界气象组织准备使用 BUFR(二进制通用数据表示格式)码。此格式具有灵活、简洁和自定义的特性。但结合我国的具体情况,BUFR 在我国广泛推广使用还需要一段时间。鉴于这种情况,我们重新编制了适合目前易于全国推广的标准格式。

在上述格式设计中,我们对一些问题做

了专门的技术处理,力求格式设计科学、合理、实用。其中主要有:

1)在格式设计中,对某些不固定长记录采用了索引技术进行了处理。由于不定长记录每次所需存储空间是随机、不固定的,而在给各个不定长分配存储空间时,需留出最大的存储空间。这样就会造成很大的空间浪费。采用索引技术,在格式中设计索引表,表中存放各个不定长记录相对首地址和数据个数,并把不定长记录集中放在格式的后面,统一留出空间,这样各个不定长记录可在这个空间中调配使用,大大节省存储空间。由于采用了索引技术,在使用时可按固定长记录使用。对不定长记录只要从索引中就可直接找到相对首地址和数据个数。这样,方便了使用并大大地节省使用时间。

2)为了说明资料质量情况,使用户使用时做到对资料的质量好坏心中有数。对此,在格式设计中专门设置了质量控制项,用来专门说明质量情况。为了区分质量的好坏,设计了质量控制码,例如,控制码为1说明已做质量控制,该值正确,控制码为2,说明未做质量控制,该值被怀疑有错等等。

3)由于国内地面测站各要素的观测方式和信息化方式的不一样,导致了记录格式的多样化。为了区分各要素各种不同的记录格式,在格式中设置了要素特征方式位,这样大大地方便了使用。

目前,国家气象中心资料室已根据制定的标准格式,对全球陆地测站地面气象资料、全球海洋测站地面气象资料、国内高空测站资料、国内陆地测站地面气象资料均已进行格式转换,并已投入业务使用,认为格式设计合理、使用方便。

3.2 气候资料数据压缩方法的研究

压缩技术是一门信息处理的科学,广泛地应用于通讯、语音、图像处理、信息存储等领域。所谓压缩技术就是在不丢失信息或以一定质量损失容限的基础上,通过改变数据的表示及存储格式,简单地说,就是启用有效

的编码来表示数据,使得数据存储时,能够减少存储空间,在通讯传输时,能够缩短传输时间,减少传输带宽。

气候资料信息量如此之大,必须进行数据压缩,以减少存储空间,缩短传输时间。数据压缩方法是多种多样的。我们对各种数据压缩方法进行了认真的分析研究,并进行了试验。根据研究结果,我们针对不同类型资料的特点,分别采用不同的压缩方法。目前我们采用如下一些方法:

1)半字长存储压缩方法

所谓半字长存储压缩方法,就是每个数值占半个字(二个字节),这样,它可容纳-32768到+32767或0~65535范围内的任一数值,而这个数值范围,对一般气候原始数据来说,已远远满足了。这种压缩方法,我们主要用于国内地面基本站和国内高空站资料数据压缩存储。因为这类资料在气候资料服务中,使用单位最多使用频率也最高。用户使用这种压缩存储数据时,不需进行任何解码工作,就可直接进行加工统计,这样就可大大地节省使用时的解码工作。

2)二进制压缩码

BUFR码和GRIB码都是世界气象组织(WMO)建议推广使用的两种具有高压缩率的二进制代码。GRIB码适用于表示数值天气预报系统加工分析和预报产品;BUFR码适用于由GTS通讯线路传来的各种观测资料,它们都具有简洁、灵活和自定义性,并且都与计算机类型无关,压缩率一般在50%以上。正是这种高压缩率使得传输速率相对地提高了,并减少了磁盘、磁带存储空间。BUFR码的自定义特征指的是一份BUFR报中,不但包括了气象数据,还包含这些数据的完整描述:有关数据物理含义、单位、精度、压缩方法及数据所占的比特位数等。这些信息即“数据描述”都包含在表中。这些表是BUFR文件的主要内容。正是这种自定义性,使BUFR码具有很强的适应性,如果要发展一种新的观测或观测平台,则只要在

BUFR 文件中增加新的观测所涉及的要素“数据描述”即可,不需要另一套电码来表示和传输这种新资料(这种压缩方法详情参见北京气象中心文集第 8901G 期)。

这种压缩方法虽具有很多的优点,尤其对永久保存的数据,因压缩率高,可大大减少存储空间,其优点更为突出。但根据我们对这些压缩数据使用的情况来看,使用起来很不方便。例如,为了寻找某一报类某时次的资料,就得把整个报类进行解码。这样,对经常使用的一些数据,使用时需将整个报类反复进行解码寻找。根据这种情况,为了弥补这种缺陷,我们已经开始把一些经常使用的要素数据从天气报资料中提取出来,再利用半字节存储压缩方法,把数据存储到数据库中或其它载体上,以便提供服务使用。

4 气候资料数据存储系统的研究

数据存储系统是数据管理系统的重要组成部分,是记录数据的载体。目前,我国气候资料数据存储方式与先进国家相比,还是非常落后的。作为气候资料数据存档及应用服务的载体,主要是 6250BPI 磁带。这样,一方面不利于数据永久保存,另一方面大量的数据产品由于缺乏可靠、快速存储系统,不可能高效地提供服务。因此,建立高效快速存储系统已迫在眉睫。尤其是随着 9210 工程的开展,大量实时资料需要存储。因此建立气候资料数据存储系统意义更为重要。

近年来,我们对气候资料数据存储系统进行了调查和研究。根据需要与可能,我们认为国家级和区域中心应建立以硬盘、光盘和磁带组成的多级存储系统。当前和今后一定时期内,直接快速动态存取(与高速缓冲存储联用)需使用硬盘,作为一级存储系统,主要存储服务频率高和业务系统中经常使用的数据。由可重写的光盘堆或光盘自动换盘机组成的超大容量光盘存储载体作为二级存储系统。磁带或自动磁带库是联机的第三级存储系统。省级应建立硬盘-光盘(数据流)二级存

储系统。

5 气候资料数据检索方法的研究

气候资料数据检索是气候资料数据管理系统中的重要组成部分,是面向用户的最终环节,它直接影响到气候资料数据服务的时效。因此,我们针对气候资料数据存储的不同载体和管理方式,研究不同的检索方法,以提高检索速度和服务时效。

5.1 磁带库检索系统的研究

随着时间的推移,科学探测手段的发展,气象资料信息量与日俱增。目前,气象资料信息基本载体为磁带。面对这样庞大的信息量,如何对磁带信息进行科学管理、存储和使用已成为气候资料工作者面临的需迫切解决的问题。为了加强气象资料的信息管理,更有效地挖掘气候资源的潜力,充分发挥气象资料的经济效益,及时准确、方便可靠地提供气象资料信息服务,在当前条件下,建立磁带库检索系统是适宜而十分必要的。

随着商用数据库管理系统(DBMS)的迅速发展,有可能建立大型气候资料数据库,但仍需以磁带库检索系统为基础后援。虽然,随着计算机技术迅猛发展,获得越来越大的硬盘是较容易的,但计算机存储容量总是有限的,气候资料数据库不可能容纳所有的气候资料信息。再则,气候资料的使用具有很大的随机性,把大量的并非经常使用的数据存入数据库中会造成很大的设备浪费,是不合适的。可以说,在一定的时期内,数据库不可能完全代替磁带库检索系统,建立和使用磁带库检索系统仍有很大的实际意义。

多年来,我们就如何建立磁带库检索系统进行了研究和尝试。我们着重对磁带库检索系统的结构、功能和组成部分以及检索工作方式、途径进行了研究。

不同的计算机,不同的资料种类配备的功能模块会有差别,但对任何气候资料磁带库检索系统来说,其逻辑结构和主要功能模块应是相同的。详情见图 1 和图 2。

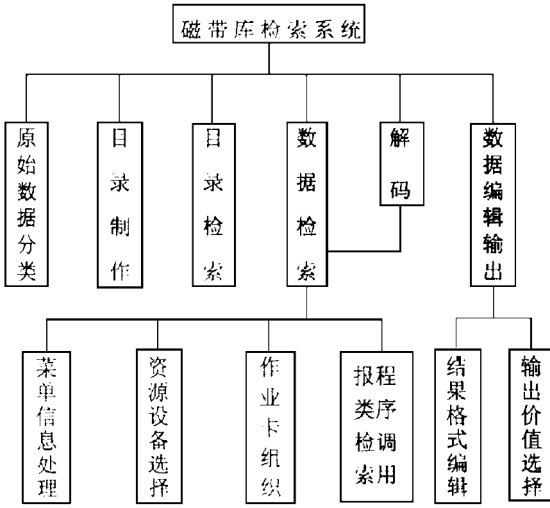


图1 系统功能层次示意图

国家气象中心气候应用室是全国气象部门中存储和管理气候资料磁带的主要单位。建立磁带库检索系统是必不可少的。结合“七五”、“八五”的科技攻关,已先后在 M-360 计算机上建立了以“全球常规气象观测资料”和“要素库、格点场库压缩资料”为对象两个磁带库检索系统,1980 年以来由实时资料磁带存档的所有历史气象资料都能得到方便使用。

5.2 气候资料光盘数据检索方法的研究

光盘技术是 70 年代发展起来的大容量

数据信息存储技术,它是 80 年代世界电子科技十大开发项目之一,是 90 年代计算机外设新型产品。光盘存储具有容量大、密度高、抗污染、位价格低、介质寿命长等特点。虽然它问世只有二十年的历史,但由于可作为数据、文字、图像、声音等多媒体信息载体。因此,在许多行业中得到应用,并起着越来越重要的作用。众所周知,气候资料具有类别多、数据量大、增长率高、服务面广以及用途千差万别等特点。因此,光盘技术作为气候观测记录存储载体,具有广阔的应用前景。为了在气象部门能尽快的启用光盘技术,在“八五”期间,我们对气候资料光盘数据的存储及光盘数据的检索技术进行了研究,以便确定气候资料光盘存储数据结构和寻求数据检索方法,并以国内地面基本站 30 年气候标准值为试验数据,建立了国内地面基本站标准值光盘数据系统,系统的逻辑结构见图 3,功能结构见图 4。

对气候资料光盘数据检索系统的研究表明:

1)光盘技术非常适于气候资料数据的存储和加工处理,尤其是光盘文档管理系统的开发和应用,使气象图、文档的高密度存储、快速检索利用成为现实。

2)光盘存储数据的数据文件结构和使用

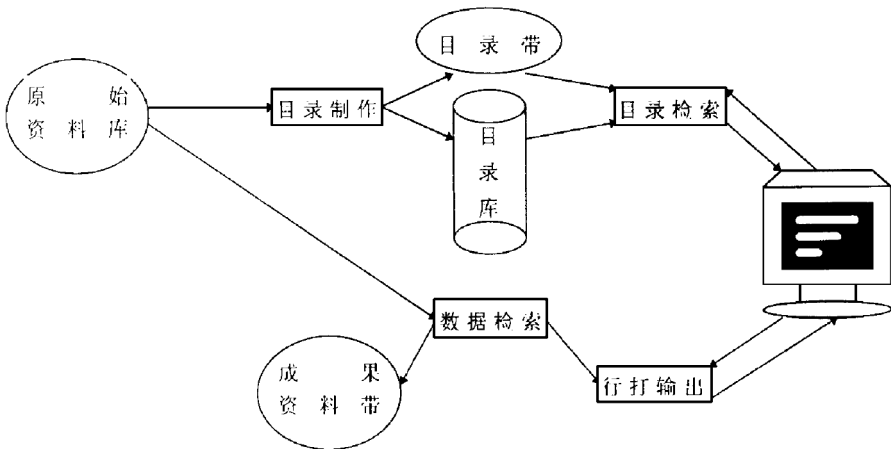


图2 磁带库检索系统逻辑结构图

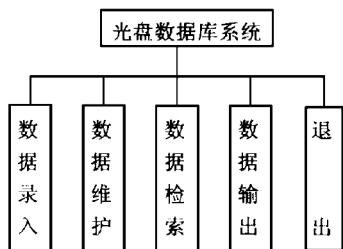
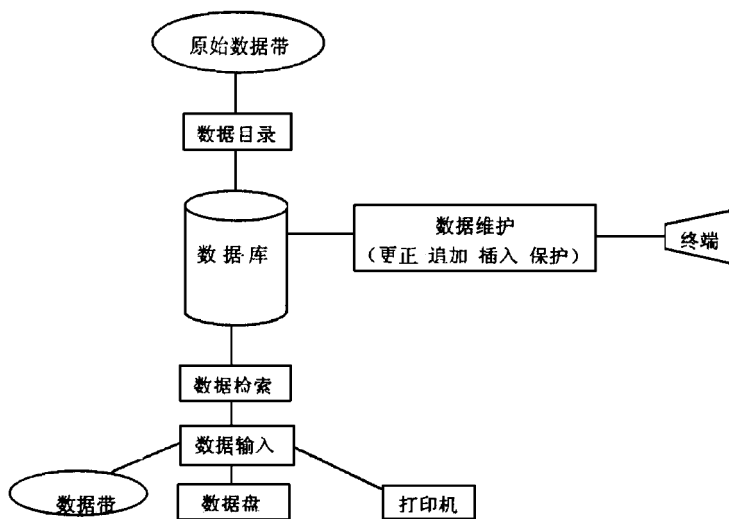


图 4 光盘数据库系统功能示意图

方式,完全和硬盘的数据文件结构和使用方式一样,简易方便。

3)经论证后,光盘技术已在许多省资料部门得到推广应用,实践证明是可行的。

5.3 分布式数据库技术在气候资料中应用的研究^[1]

分布式数据库技术是新一代数据库技术,它是在集中式数据库的基础上,由计算机网络的引入而发展起来的新技术。所谓分布式数据库从逻辑上看是一个完整的数据库,而实际上它的物理数据库是分布在计算机网络的不同节点上。网络中每个节点具有独立处理本地数据库的功能,可执行局部应用,同时也可存取和处理多个异地数据库的数据,执行全局应用。

在“八五”期间,中国气象局投资建设“气象资料处理与分析服务系统”工程项目中,建立了客户/服务器(Client/Server)结构的计算机环境,并选用了最优价格性能比的 Sybase 分布式数据管理系统,为这项研究提供了很好的软、硬平台。我们以 SUN670, PCS(微机服务器)构成双数据库服务器,模拟分布处理环境,并结合实际工作需要,将全国国家基本站的部分要素的逐日值放在 SUN670 服务器上,标准值和月值放在微机服务器(PCS)上,这样就构成了国家基本站气候资料分布式数据库系统。其功能见图 5。

5.3.1 系统管理功能

①数据录入功能:建库时通过批量数据入库控制程序(BCP)进行批量数据入库,后续资料的追加,通过数据库(DB)—文件库(LIBRARY)编的与 A₀、A₁ 文件的接口程序,与实时库的接口程序进行数据追加。

②用户管理:通过客户机上的数据库管理员(DBA)应用程序进行增加、删除用户,授予收回查询表,执行存储过程的权利。

③完整性控制:在数据库表中建立适当的规则,以保证表内数据的完整性,建立相应于插入、删除等的触发器来保证表之间参照

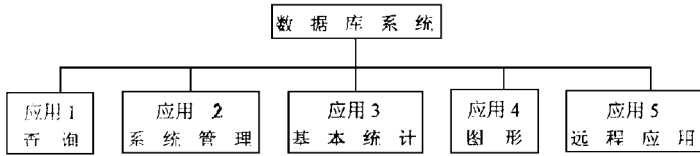


图5 国家基本站气候资料分布式数据库系统功能图

的完整性。

④备份恢复功能

5.3.2 查询功能

查询是数据库应用系统中最重要也是最常用的功能。

①标准查询:提供给用户一种比较标准的数据查询方法,用户按要素、时间、台站、省份和台站经纬度进行查询。由于标准查询语言(SOL)本身的局限以及对存储过程的限制,在标准查询中,用户一次只能查询一种要素、一个时间或时段,一个省或少于20个台站的数据。由此可见,标准查询方式适用于规模不大,比较标准的用户需求,对于需要更加灵活可变查询手段的用户,就需要用下面的动态查询。

②动态查询:在系统开发过程中,经试验可采用Powerbuilder工具中的动态SQL实现动态查询。动态SQL语句,就是在非运行时查询语句的各个部分还不确定,只是在运行过程中才一步步装配起来的SQL语句。由于存储过程不支持动态SQL,所以我们把SQL语句的生成、编译移到客户机的应用程序一侧。通过动态SQL语句,用户即可以灵活、随机的获取满足自己条件的气候资料。由以上描述可知,标准查询和动态查询实现的内在机理是不一样的,但对用户来讲,是透明的,两者的窗口风格、操作方式都基本一致。

5.3.3 基本统计分析模块和图形模块

系统可以在库内数据的基础上进行一些基本的气候资料数据的统计分析。统计值可以图形表示,如求某一时段数据的均值、方差、极值、滑动平均等。其值可用线性图、柱形图、饼图等不同的形式表示出来。

通过对国家基本站气候资料分布式数据库系统的研究,有下列一些体会:

①研究表明,利用商用分布式数据库管理系统建立气候资料分布式数据库是可行的。

②我们认为建立分布式数据库比集中式数据库具有更大的可靠性和可用性,这是由于分布式数据库系统是多位置上由多台计算机构成的,在个别节点或个别通讯线路发生故障的情况下,它仍然可以继续工作。

③提高效率,降低了通讯费用。这是由于分布式数据库都是把数据放在经常使用的节点上,这样做不仅减少了系统的响应时间,提高了效率,而且也降低了通讯费用。

④增强了数据库容量的可扩充性。这是由于在已有的分布式数据库中增加一个新节点,比用一个更大的系统代替一个已有的集中式数据库容易得多和节省得多。

⑤气候资料分布更新。我们认为使用复制服务器方式较为适合,而不使用两阶段提交方式。复制服务器方式是一种松耦合模式的分布更新方式,它允许主数据节点与其它节点数据有一定长时段的 inconsistency。在对数据时间性要求不很高的情况下,这是一种比较适合的气候资料分布更新方式。两阶段提交分布更新方式是一种紧耦合模式的分布更新方式,它可以保证主数据节点与其它节点数据实时的一致性,对时间要求很严,故不适合气候资料的分布更新。

6 结语

多年来,我们虽然对气候资料数据管理技术进行了有关方面的研究,取得了一定成

绩,但与发达国家相比,还有很大差距。在资料存储方式、信息网络系统以及数据管理方式等方面都比较落后,并且尚未形成完整的气候资料数据管理业务系统。因此,我们认为目前首先要对气候资料数据管理业务系统的结构、功能以及相互之间接口进行研究,逐步形成以分布式数据库为中心的高度集成的气

候资料数据管理业务体系结构,以适应服务的需要,使气候资料工作现代化建设上一个新台阶。

参考文献

- 1 宋连春等. 分布式数据在气候资料管理中的应用研究. 国家气象中心科技年报, 1996