

宁夏综合气象信息共享与管理系统设计研究

李新庆^{1,2} 陈海波^{1,2} 杨有林¹ 卫建国¹ 马宁¹ 陈增境¹

(1. 中国气象局旱区特色农业气象灾害监测预警与风险管理重点实验室, 宁夏银川 750002;

2. 宁夏气象防灾减灾重点实验室, 宁夏银川 750002)

摘要: 气象数据具有多源、异构、海量、高时效等特点, 如何便捷高效地提供共享管理是一个长期困扰业务和管理人员的难题。宁夏综合气象信息共享与管理系统着力解决气象数据集约化、标准化、智能化的收集、存储、共享和可视化展示等问题, 为提升气象数据共享和信息管理服务水平提供了很好的实践示范。文中首先描述该系统的总体架构, 阐明了系统各项功能模块的设计思路; 其次, 基于 CIMISS 和 Hadoop HDFS 构建的省(区)级数据存储环境, 提出了基础数据库、业务产品库和服务产品库的逻辑划分, 解决了 CIMISS 对非结构化数据存储效率不足的问题; 再次, 描述了系统所涉及的个性化推荐、智能搜索、数据可视化等核心应用技术, 为数据共享管理提供了更多“智慧元素”; 最后, 给出了系统的性能测试结果和应用效果, 表明该系统能够为宁夏气象业务用户提供高效便捷的数据共享和信息管理服务, 具有一定的示范推广价值。

关键词: 气象数据; 信息共享; CIMISS; Hadoop; 个性化推荐; 智能搜索

中图分类号: P49; P409

文献标识码: A

文章编号: 1673-629X(2019)05-0135-07

doi: 10.3969/j.issn.1673-629X.2019.05.029

Design and Research of Ningxia Integrated Meteorological Information Sharing and Management System

LI Xin-qing^{1,2}, CHEN Hai-bo^{1,2}, YANG You-lin¹, WEI Jian-guo¹, MA Ning¹, CHEN Zeng-jing¹

(1. Key Laboratory for Meteorological Disaster Monitoring and Early Warning and Risk Management of

Characteristic Agriculture in Arid Regions, CMA, Yinchuan 750002, China;

2. Ningxia Key Lab of Meteorological Disaster Prevention and Reduction, Yinchuan 750002, China)

Abstract: Meteorological data is heterogeneous and massive with multi-source and high timeliness. How to provide sharing management conveniently and efficiently is a difficulty that has been troubling business and managers for a long time. Therefore, Ningxia integrated meteorological information sharing and management system focuses on solving the problems of meteorological data collection, storage, sharing and visual display, which are intensive, standardized and intelligent. It provides great practice demonstration for improving meteorological data sharing and information management service level. Firstly, we expound the overall structure of system and explain the design ideas of various functional modules. Secondly, based on the provincial (autonomous region) data storage environment built by CIMISS and Hadoop HDFS, we propose the logical division of basic database, business product database and service product database, which solves the problem of CIMISS's low efficiency in unstructured data storage. Thirdly, we describe the core application technologies including personalized recommendation, intelligent search and data visualization in the system, providing more "intelligent elements" for data sharing management. Lastly, the system performance testing and application effect are briefly described, which shows that the system can provide data sharing and information management services for the whole region meteorological service users, with a certain value of demonstration and promotion.

Key words: meteorological data; information sharing; CIMISS; Hadoop; personalized recommendation; intelligent search

收稿日期: 2018-06-11

修回日期: 2018-10-12

网络出版时间: 2018-12-21

基金项目: 国家自然科学基金(41565007); 宁夏自然科学基金(NZ16202); 中国气象局旱区特色农业气象灾害监测预警与风险管理重点实验室开放基金(CAMF-201817)

作者简介: 李新庆(1983-), 男, 硕士, 工程师, 研究方向为气象数据应用; 陈海波, 硕士, 高级工程师, 通信作者, 研究方向为气象数据应用。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20181221.1554.044.html>

1 概述

随着气象信息化、现代化事业的不断发展,宁夏回族自治区气象业务信息化水平得到显著提高,气象业务系统建设也得到大幅度的提升,近些年自治区级气象业务单位自建的重要业务系统 46 个,系统涉及综合观测、天气预报、气候预测、农业气象、人工影响天气以及公共气象服务等业务体系,并在实际业务工作中发挥了一定的效益。

但与此同时,在缺乏整体规划的背景下,职能处室、业务单位之间缺乏沟通,闷头自我发展成为历史常态^[1],系统建设也暴露出了一些突出问题:业务系统建设整体规划不够,综合统筹乏力,建设各自为政,造成系统设计随意化、开发个人化、功能碎片化;在建系统“小而散”现象突出,各系统业务流程冗杂,应用效率较低;各系统中数据格式、数据流向复杂,数据反复搬家,造成“信息孤岛”遍布、“应用烟囱”林立的局面。

打破数据壁垒、整合信息资源、简化流程环节、推进业务融合与协同^[2],促进气象数据共享和业务科学管理已经成为宁夏气象业务发展的亟待解决的关键问题。

宁夏回族自治区气象局于 2016 年启动智能化综合气象业务服务管理共享平台“331”建设任务,即以全国综合气象信息共享平台(China integrated meteorological information sharing system, CIMISS)环境构建宁夏气象综合数据库,包括基础数据库、业务产品库、服务产品库 3 个数据库;建设业务产品生成系统、服务产品制作系统和综合发布系统 3 类系统;建设 1 个综合气象信息共享与管理系统。其中,综合气象信息共享与管理系统旨在整合宁夏气象内部数据资源,为气象行业用户提供一个包括实况观测资料、历史气候资料以及管理信息等功能于一体的信息共享与管理系统,实现天气预报、气候预测、农业气象、人影天气、专业服务等核心业务的在线共享服务,同时为气象业务管理提供辅助决策支持。

建立海量气象数据的管理和服务系统是气象业务和科学研究的先导性工作^[3],同时也是气象信息化发展所面临的难点工作之一。因此,近年来国家、各省级气象部门专家纷纷开展了气象信息共享系统的研究和建设工作,并取得一定的进展。文献[3-4]提出全国综合气象信息共享平台(CIMISS)的设计与实现,初步建成统一、标准的基础气象数据库,包括实时观测数据、历史数据和业务产品等 14 大类,498 个子类,1 500 个细目的气象资料集中存储,为国家级和省级气象行业用户提供标准一致的数据环境,在一定程度上缓解了气象数据共享不足的问题;何林等^[5]依托 CIMISS 系统研发了陕西省气象数据共享系统,实现了 CIMISS

所接入的地面、高空、农气、辐射、雷达、卫星等 14 类资料的查询下载、统计分析、可视化展示等功能,并针对 CIMISS 中缺少的本省特色数据进行适当补充,为省级用户提供完整、及时、稳定、准确的对内气象数据服务;王宏记等^[6]提出了基于 CIMISS 的长江流域气象水文信息共享系统的设计思路,研究了信息收集与处理、数据存储管理和信息共享服务等关键技术问题,梳理了流域内监测产品、预报预测、灾情和应急服务等综合信息,初步建立了具有流域特色的水文气象信息共享规范。除此之外,季刚^[7]、张立^[8]、杨伟明^[9]、史彩霞^[10]等均开展了省级综合气象信息共享系统的研究与应用,均取得了一定进展。从上述整体业务应用情况归纳来看,基于 CIMISS 统一数据环境构建多方面业务应用,已经成为各地支撑气象业务和服务系统的主要形式。

为实现气象数据资源跨部门的充分流通与共享,有效解决省际气象数据汇交、标准化存储、数据可视化等问题,文中提出了宁夏综合气象信息共享与管理系统设计与实现的关键技术方法。首先,对宁夏综合气象信息共享与管理系统的总体框架设计和各功能模块进行阐述。其次,基于 CIMISS 环境对数据存储进行扩展,构建宁夏气象综合数据库,实现省(区)级气象数据的集中存储,为共享与管理系统提供数据支撑。再次,为提高数据服务,提出了基于内容相似度的个性化推荐算法实现气象资料的智能筛选,为用户提供更为个性化的数据推荐服务;然后,提出了一种基于 ElasticSearch 分布式搜索引擎的气象数据智能检索技术,为用户提供快速、高效、可靠的气象数据检索和共享服务。最后,结合 HTML5、Web GIS 等 Web 技术,构建了宁夏气象综合信息共享与管理系统,实现了多源气象信息的集成展示,进一步提升了宁夏气象信息共享服务水平。

2 系统设计

2.1 总体架构

宁夏综合气象信息共享与管理系统致力于在气象行业内部满足各业务信息系统之间的数据资源共享和交换,为省(区)、市、县三级气象业务和管理用户提供信息共享和一站式在线服务。归纳起来,共享信息主要包括基础数据、业务产品、服务产品和管理信息数据。

其中,基础数据是指由观测设备或国家气象部门广播上传、下发的各类气象基础数据;业务产品主要是指由核心气象业务系统在运行中,通过模型分析和计算所产生的统计值、图形等中间加工产品;服务产品主要是指面向决策服务、防灾减灾、公众服务等专业系统

所生成的最终服务产品。

实现上述信息的统一共享和管理需求,系统必须制定标准、规范的数据存储格式,实现多源、异构数据的集中存储和共享,功能上包括数据资源目录管理、元数据管理、共享信息存储管理、系统资源管理、数据服务接口管理和系统监控等诸多信息^[11]。为此,在遵循 CIMISS 数据标准和规范下,参考气象资料和气象要素分类编码标准,对宁夏各部门数据资源进行梳理,在

CIMISS 数据库的基础上构建了宁夏气象综合数据库,在逻辑上划分为基础数据库、业务产品库和服务产品库,实现对结构化数据、非结构化数据的集约、高效存储。同时,以虚拟化资源池为基础,采用“云+端”模式建立信息共享与管理系统,使其恰好满足气象业务“弹性扩展和计算”的需求^[12],为系统提供快速扩展、可定制的气象数据加工处理和共享服务,以解决“信息孤岛”、“烟囱林立”的业务碎片化局面。



图1 综合气象信息共享与管理系统总体架构

系统总体架构如图1所示,采用成熟的三层体系架构^[13]:表示层、业务逻辑层和数据层。表示层是系统门户,采用标准的Web客户端开发,负责与用户进行交互,提供数据智能检索、快速导航、个性化推荐和数据可视化展示等功能,通过Web GIS、图形、表格等方式实现各类气象数据的共享服务。业务逻辑层除了提供基础数据、业务产品、服务产品和管理信息的接口访问,还提供对用户行为、日志信息等系统本身的管理服务。数据层主要依赖 CIMISS 数据环境构建宁夏气象综合数据库,实现对实时数据、历史数据和管理信息的集中存储管理,通过数据服务接口响应业务层的数据请求,为系统提供数据支撑。

2.2 功能设计

宁夏综合气象信息共享与管理系统包括基础数据、业务产品、服务产品和管理信息共四个功能模块。

(1) 首页。首页作为系统的访问入口,综合展示气象基础数据、业务、服务和信息管理等重点关注的信息,功能除了包括各类导航、通知信息外,还包括智能

检索、个性化推荐、最新产品链接等。

(2) 基础数据。基础数据通常是行业用户需求最大的一类气象科学数据,根据气象资料的内容属性和 CIMISS 分类标准,功能分类包括地面、高空、辐射、农气生态、大气成分、大气环流及海洋、雷达、卫星、数值预报产品和其他共10类气象资料。除了包括查询、统计和下载等常规功能外,还包括如时序图、分布图、等值线、色斑图的在线制图服务。

(3) 业务产品。业务产品主要面向气象业务、科研用户,功能覆盖气象部门核心业务单位,包括天气类、气候类、农气类、专业服务类和人影类的业务产品。除提供常规查询功能外,还提供加工统计值、分布图等在线分析功能。

(4) 服务产品。服务产品主要面向决策、公众和专业服务,将气象基础数据、服务产品加工制作和产品分发有机融合,提供服务产品的可视化展示和下载功能。

(5) 管理信息。管理信息面向气象管理部门用

户,实现气象业务、事务管理数据的规范化、扁平化,为气象管理提供辅助决策支持,提升执行力和预判力。功能上主要包括数据监控、质量监控和信息通报。数据监控包括基础数据、业务产品、服务产品从收集、解析、入库到共享服务一体化、全流程的监控管理^[14];质量监控包括传输质量(CIMISS、MDOS)、预报质量、预警质量等信息的监控;信息通报包括现代化建设、产品发布、数据下载通报等。

3 系统关键技术

3.1 海量异构气象数据存储

气象业务领域内每天的数据增量以 TB 计算^[15],气象数据对天气预报、气候研究、公共气象服务以及各个行业的应用都具有较为重要的价值,需要将各类数据按照统一的标准进行存储和管理。CIMISS 作为国省级的数据中心,为气象行业用户解决了气象数据收集、加工处理、数据存储的标准化、集约化等若干难点问题,但是 CIMISS 采用关系型数据库存储非结构化数据的元数据信息、采用集群文件系统存储非结构化数据文件,上述传统架构对不断增加的海量非结构化气象数据存储服务功能不足^[16],易出现负载饱和、读写性能不足的瓶颈问题,影响了系统服务的时效性和稳定性。

因此,对 CIMISS 数据存储进行扩展应用研究,提出基于 Oracle 数据库集群和 Hadoop 分布式文件系统混合架构的气象综合数据库,实现省(区)级气象基础数据、业务产品、服务产品和管理信息的数据、标准、统一、集约存储环境。

结构化数据作为共享和服务应用范围最广,访问最频繁的气象数据^[17],必须保证系统对结构化数据的快速处理和存储能力。

因此,该类数据以 CIMISS 数据源为基础,数据标准保持与 CIMISS 数据环境的一致性,有利于本区域新增数据和气象标准统一,达到现有业务与原有业务的有机融合。

对于大量非结构化文件的快速处理和存储是信息共享和服务的关键,特别是以雷达、卫星为代表的大文件和报文为代表的海量小文件^[18],需要的耗时和占用的资源问题较为突出。为此,文献[19-22]针对文件存储都作了大量研究工作,有效解决了非结构化数据存储的问题。文中针对上述数据采用 Hadoop HDFS 分布式文件存储,总容量规划为 10 TB,设计了两个数据节点(DataNode),实现了非结构化数据的集中存储。

气象综合数据库采用客户端、Web Service 和 REST(representational state transfer)三种服务接口方

式对外提供数据服务,提供基础数据服务接口 135 个、业务产品服务接口 452 个、服务产品服务接口 358 个,实现了入库数据和产品的全覆盖。

3.2 气象数据加工处理

气象数据加工处理对收集的观测数据进行数据抽取、数据融合、统计加工处理,根据用户制定任务模板自动生成统计产品和图形化产品,最终提交至气象综合数据库中进行存储管理和共享服务。其核心依赖于气象常用算法模型库(常用统计算法、差值算法和图形化加工算法),按照上述算法实时、准实时和定时地生成各类加工产品。

加工处理系统按照业务功能分为常规统计产品加工、网格化产品加工和图形化产品加工三类业务产品。常规统计产品以中国气象局《地面气象观测规范》、《地面气象资料实时统计处理业务规定(2017 版)》为标准,功能包括日、候、旬、月、季、年、30 年气候统计值共 2 126 种气象统计要素;网格化产品是以中国气象局 5×5 公里实况格点化产品为基础,融合宁夏区域小时、日等地面观测数据,细化生成 1×1 公里的本地特色气象要素格点化产品;图形化产品以宁夏本地的图形化产品标准,实时、准实时地生成落区图、分布图等 132 种图形化产品。

数据加工处理流程包括以下步骤:任务调度、数据加工、读取接口、算法模型和写入模块。任务调度采用 Quartz 开源组件实现,Quartz 是一个轻量级的企业调度框架,完全由 Java 编写,其提供了强大的任务调度机制,同时保持了使用的简单性^[23]。加工处理系统首先通过 Quartz 任务调度框架定时触发加工任务,加工任务再调用服务接口获取任务要求的结果集;然后将结果集以及数据模型的统计参数传入到算法模型,加工流程再根据算法模型返回统计或运算完成的结果集;最后将运算结果传送到接口进行数据的写入。数据加工处理中每个步骤都会进行日志记录,为以后日志分析提供参考依据。

3.3 个性化推荐

为解决海量气象数据服务问题,采用个性化推荐功能提高用户访问和获取数据的服务能力。个性化推荐已经作为处理“信息超载”问题的主流方法^[24],其优势在于主动收集用户特征资料,根据用户的兴趣特征,从海量信息中挖掘用户可能感兴趣或需要的信息资源,对用户做出相应推荐。

基于用户的个性化推荐算法涉及到用户、产品之间的相似度计算^[25],利用大量用户和产品关联的历史数据,采用基于余弦相似度算法来计算用户和产品之间的相似度,查找与目标用户相似性较高的邻近值,通过邻近用户集对其产品的评分来预测用户对该产品的

评分,得到智能推荐的产品集合,为气象行业用户提供个性化推荐服务。事实上,一个完整的推荐系统应该由用户模型、产品模型和推荐算法共同组成,具体算法定义如下^[26]:

首先,收集用户行为,设 $U = (U_1, U_2, \dots, U_n)$ 为用户集合, $G = (G_1, G_2, \dots, G_m)$ 为给定产品集合,生成用户对产品的评分矩为 $R_{n \times m}$ 。其中, n 行表示共有 n 个用户, m 列表示共有 m 个产品, $R_{n \times m}$ 表示第 n 个用户对第 m 个产品的评分。

其次,建立用户和邻近对象的相似度,把用户 u 和

$$\text{sim}(u, v) = \frac{\sum_{i \in N(u) \cap N(v)} (R_{u,i} - \bar{R}_u) \times (R_{v,i} - \bar{R}_v)}{\sqrt{\sum_{i \in N(u) \cap N(v)} (R_{u,i} - \bar{R}_u)^2} \times \sqrt{\sum_{i \in N(u) \cap N(v)} (R_{v,i} - \bar{R}_v)^2}} \quad (2)$$

其中, \bar{R}_u 表示用户 u 对已评价推荐对象的平均评分; $R_{u,i}$ 表示用户 u 对推荐对象 i 的评分。

再次,生成邻近用户。根据计算目标用户和其他用户的相似度后,按照大小顺序排列,得出前 k 个最大的邻居组成目标用户的邻近 k 个子集。

最后,产生推荐结果。利用公式计算出用户对产品的预测分值。

$$P_{u,i} = \bar{R}_u + \frac{\sum_{v \in S(u)} \text{sim}(u, v) \times (R_{v,i} - \bar{R}_v)}{\sum_{v \in S(u)} (|\text{sim}(u, v)|)} \quad (3)$$

其中, $P_{u,i}$ 为用户 u 对产品 i 的评分值; $S(u)$ 表示用户 u 的邻居集合; \bar{R}_u 和 \bar{R}_v 分别为用户 u 和邻近用户 v 对已评分产品的平均评分。

上述算法复杂性不高,且能够较好地挖掘气象数据及相关产品,为气象行业用户提供个性化数据和产品需求。

3.4 智能搜索引擎

传统的数据检索方法已经难以满足日益增长的气象数据高性能检索需求,如何快速、高效地从海量气象数据中检索有用的数据资源已经成为共享与管理系统的难点之一。

因此,结合个性化推荐和分布式搜索引擎,提出了基于 ElasticSearch 构建智能搜索引擎,针对海量气象数据实现自动检索相关信息,同时为用户提供快速信息查询服务。

ElasticSearch 是一个基于 Lucene 开源检索引擎,具有多租户支持、分布式、高可用性、全文搜索等特征,可以方便地嵌入到各种应用中实现对应用的全文索引和检索功能^[28]。ElasticSearch 全文检索引擎采用 RESTful 接口,提供了简单易用的查询和共享接口。ElasticSearch 可以非常简单地实现分布式搜索,自己带有分布式管理模块,可以方便、迅速地进行集群的横

用户 v 对 (G_1, G_2, \dots, G_m) 中的评分记作 R_u 和 R_v 两个评分向量,那么,用户 u 和用户 v 的相似度等于 R_u 和 R_v 的余弦夹角。则 $\text{sim}(u, v)$ 的相似性计算方法如下:

$$\text{sim}(u, v) = \cos(R_u, R_v) = \frac{R_u \times R_v}{\|R_u\| \times \|R_v\|} \quad (1)$$

由于在余弦相似度度量方法中没有考虑不同用户评分尺度的问题^[27],根据 Pearson 修正的余弦相似度度量方法,通过减去用户对项目的平均评分来改善上述问题。

向扩展并自动化进行集群的负载均衡,从而高效、稳定地对 PB 级海量气象数据进行检索和分析。ElasticSearch 能够方便地建立索引,将一个索引分割成多个索引分片,然后将多个分片均衡分布在集群的所有节点上。

构建 ElasticSearch 搜索引擎包括分布式集群设计、垂直领域专业词典构建、数据索引、数据搜索和个性化推荐。ElasticSearch 分布式集群根据现有宁夏气象基础数据、各类产品规模及相关性能进行集群设计,包括集群各节点设计、集群相关配置。在垂直领域专业词典构建中,根据每一个垂直领域建立一个对应的个性化词典。当检索的数据或产品涉及到对应的垂直领域时,就把对应的个性化词典加入到分词词库中。虽然 ElasticSearch 内置了多种分词器,但是对中文处理效果不够理想^[29],因此,文中采用 jieba 分词组件,该组件是目前最好的 python 类中文分词器^[30],支持精简模式、全模式和搜索引擎模式进行分词。个性化词典库的建立主要分为两部分,一部分是基于气象元数据建立通用词典,根据梳理的气象元数据,建立 5 832 个关键词;另一部分是使用 GET 方法来定期对检索数据的新词进行发现,并将该新词存储到个性化词库中。当使用 ElasticSearch 进行索引和搜索操作时,针对索引和搜索涉及的垂直领域不同,将分词组件加载到不同的垂直领域个性化词典,保证分词处理更加高效、精准。

3.5 面向气象数据的可视化技术

可视化技术融合了信息可视化、科学可视化、数据挖掘、人机交互等学科,针对气象数据具有多类型、高维度、弱模式等特点^[31],利用交互式界面对复杂气象数据进行分析 and 展示,同时融入认知能力和计算能力,从而为用户提供气象数据的高效获取和服务。

常用的可视化分析方法处理气象数据包括二维、三维、多维、文本、网络信息、时空信息等^[32]。针对二

维数据的展示,主要包括曲线图、折线图、散点图、饼图、柱状图等,Highcharts、FusionCharts、D3.js、Web GL等可视化工具都对上述类型提供了可靠的支持。文中采用 Highcharts 实现了气象要素在时间序列上的趋势图,直观反映了各种气象要素按照小时、日、旬、月、季、年不同时间尺度的展示效果。针对等值面图、落区图的绘制,除了借助于 GIS 引擎之外,Matplotlib 作为一款 python 编制的二维图形库^[33],提供了快速、准确的绘图能力,将其与 basemap 相结合,可提供一种绘制 GIS 地图的有效解决方案。因此,文中采用 Webservice 服务调用后台 Matplotlib 绘图引擎,实现气象 GIS 图像的绘制,最后再叠加至 Web GIS 地图上进行展示。多维数据则需要采用几何投影、图标、像素等多种方法,对其进行映射、投影和变换,从而进行降维展示,实现了雷达基本反射率、组合反射率、基本速度等图形的可视化^[34]。

表 1 系统查询接口访问测试

接口数据获取量/条	小时资料/ms	日资料/ms	月资料/ms	年资料/ms
1 000	224	133	232	304
5 000	284	278	342	385
10 000	486	619	561	581
30 000	1 329	1 633	1 448	1 457
50 000	2 987	2 673	2 758	2 248
100 000	5 298	3 923	4 898	5 187

4.2 统应用情况

系统采用 HTML5、Web GIS、Java J2EE 开发实现了省(区)级气象信息的共享与展示,该平台不仅提供 CIMISS 中常用的地面、高空、雷达等 10 类基础数据,还提供了 1×1 公里格点预报、首场透雨、K 指数预测、气候影响评价等业务和服务产品。

该系统已经在全区各级气象业务部门进行了推广使用,系统存储结构化数据 22.922 亿条,存储非结构化数据 2.31 TB,为授权用户提供基础数据 152 种、业务产品 70 种、服务产品 38 种,用户可在线浏览、查询、下载各类数据。

5 结束语

综合气象信息共享与管理系统是支持气象行业部门业务应用的数据服务平台,同时也是支撑气象领域核心业务、提升气象信息服务水平的重要基础,该系统的设计与研究将有效解决气象部门对多类型、规模大、业务逻辑复杂等数据共享和管理不足的难题。

文中的工作主要如下:针对海量数据存储效率不

4 系统测试及应用

4.1 系统性能测试

文中主要针对查询访问响应能力进行了系统性能测试,模拟 50 个用户的并发量,以气象综合数据库中的地面小时资料、日资料、月资料、年资料为测试对象,当系统查询时调用访问接口,在时间上和数量上进行了对比测试,测试结果如表 1 所示。

根据测试结果,当查询数量小于 10 000 条记录时,系统的响应时间均小于 1 s。当用户查询记录数在 30 000 条左右时,系统的平均响应时间为 1.5 s,当用户查询记录数在 50 000 条时,系统的平均响应时间 2.7 s。结果表明,在 50 个用户并发测试下,当查询记录数低于 50 000 条以下时,系统的响应时间均低于 3 s,系统查询性能达到了业务化需求。

足的问题,提出了 Oracle 集群和 Hadoop HDFS 分布式存储的混合架构的省(区)级气象综合数据库的设计实现,并对 CIMISS 数据资源管理和标准体系进行了扩展,解决了多源、异构气象数据集中存储的问题;采用余弦相似度个性化推荐和 Elasticsearch 实现了共享系统的智能搜索引擎,提高气象数据的主动检索服务;系统除了提供查询、下载各类实时、历史基础数据之外,还提供了省(区)级基础数据加工产品、中间业务产品和公共服务产品,实现了观测、预报、气候、服务等核心业务单位的数据共享全覆盖。

目前,宁夏综合气象信息共享与管理系统已经投入业务试运行,已经成为宁夏区级气象业务、服务、管理应用的重要系统之一,该系统不仅适用于宁夏气象信息共享服务,而且还可对其他省(区)级气象部门开展信息共享系统建设提供理论基础和借鉴价值。

参考文献:

- [1] 洪 蕾.“顶层设计”推动气象系统集约化发展[N].中国信息化周报,2013-12-16(06).
- [2] 马锋波,周 勇,刘东君.关于气象信息化标准体系框架编

- 制的探讨[J].标准科学,2015(21):36-40.
- [3] 熊安元,赵芳,王颖,等.全国综合气象信息共享系统的设计与实现[J].应用气象学报,2015,26(4):500-512.
- [4] ZHANG Z,ZHANG Q,XING HU,et al.Design and implementation of national meteorological service platform[J].亚洲农业研究:英文版,2016,8(9):77-81.
- [5] 何林,邓凤东,李亚丽,等.陕西省气象数据共享系统[J].陕西气象,2017(6):32-35.
- [6] 王宏记,杨代才.基于CIMISS的长江流域气象水文信息共享系统设计与实现研究[J].安徽农业科学,2014,42(32):11565-11570.
- [7] 季刚,盛绍学,江双五.安徽省气象信息共享平台[J].计算机系统应用,2013,22(12):210-214.
- [8] 张立.内蒙古气象局综合信息系统的设计与实现[D].呼和浩特:内蒙古大学,2015.
- [9] 杨伟明.基于WEB的气象信息共享平台的设计与实现[D].大连:大连理工大学,2014.
- [10] 史彩霞,黎颖智,张许斌.基于CIMISS的广西气象服务信息综合业务系统的设计与实现[J].气象研究与应用,2016,37(4):82-85.
- [11] 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[J].应用气象学报,2006,17(5):621-628.
- [12] 沈文海.从云计算看气象部门未来的信息化趋势[J].气象科技进展,2012(2):49-56.
- [13] 王涛,李强,成晓琴.基于WebGIS的祁县地质灾害预报预警系统研究[J].气象与环境科学,2017,40(1):138-143.
- [14] 王立俊,李晋峰,程洪涛,等.气象资料实时传输监控系统应用研究[J].计算机技术与发展,2018,28(3):183-186.
- [15] 沈文海.再析气象大数据及其应用[J].中国信息化,2016(1):84-96.
- [16] 季永华,孙超,刘一鸣,等.CIMISS中气象观测资料处理入库效率优化方法[J].气象科技,2017,45(1):30-35.
- [17] 黄岩.基于元数据的气象数据共享系统设计与实现[D].北京:中国科学院大学,2013.
- [18] 王若瞳,黄向东,张博,等.海量气象数据实时解析与存储系统的设计与实现[J].计算机工程与科学,2015,37(11):2045-2054.
- [19] 肖卫青,杨润芝,胡开喜,等.Hadoop在气象数据密集型处理领域中的应用[J].气象科技,2015,43(5):823-828.
- [20] 杨芙蓉,王永丽,王文明.基于Hadoop的海量气象雷达小文件存储研究[J].成都信息工程学院学报,2015,30(3):254-258.
- [21] 陈东辉,曾乐,梁中军,等.基于HBase的气象地面分钟数据分布式存储系统[J].计算机应用,2014,34(9):2617-2621.
- [22] 邹振宇,郑炅,王嵩,等.基于HDFS的云存储系统小文件优化方案[J].计算机工程,2016,42(3):34-40.
- [23] 朱哲明.基于Quartz的消息沟通平台的研究[D].北京:北京邮电大学,2012.
- [24] 刘辉,郭梦梦,潘伟强.个性化推荐系统综述[J].常州大学学报:自然科学版,2017,29(3):51-59.
- [25] 张光卫,李德毅,李鹏,等.基于云模型的协同过滤推荐算法[J].软件学报,2007,18(10):2403-2411.
- [26] 王国霞,刘贺平.个性化推荐系统综述[J].计算机工程与应用,2012,48(7):66-76.
- [27] 高鹏.推荐系统中信息相似度的研究及其应用[D].上海:上海交通大学,2013.
- [28] 孟方园,魏涛.ElasticSearch分布式搜索引擎在导航大数据检索中的应用[C]//第8届中国卫星导航学术年会论文集.上海:出版者不详,2017.
- [29] 黄翼彪.开源中文分词器的比较研究[D].郑州:郑州大学,2013.
- [30] 曾亚飞.基于Elasticsearch的分布式智能搜索引擎的研究与实现[D].重庆:重庆大学,2016.
- [31] 王建民.领域大数据应用开发与运行平台技术研究[J].软件学报,2017,28(6):1516-1528.
- [32] 任磊,杜一,马帅,等.大数据可视分析综述[J].软件学报,2014,25(9):1909-1936.
- [33] 王亚东.Python在气象数据可视化中的应用[C]//第34届中国气象学会年会论文集.郑州:中国气象学会,2017.
- [34] 王轩,路明月,郑建琴,等.雷达数据的多维可视化[J].图学学报,2017,38(4):497-503.