

基于气象大数据云平台的海南气象数据共享平台设计与实现

王双双, 杜建华, 王立俊, 赵冰, 刘骥超

(海南省气象信息中心, 海口 570203)

摘要: 为打破“数据孤岛”, 推动气象数据跨部门融合使用, 充分发挥气象数据科学价值, 针对海南省气象部门与外部门数据共享方式杂乱、共享水平低、管理滞后等问题, 依托专线网络, 从气象数据需求类型、数据格式出发, 设计了一套气象数据跨部门的分发处理流程; 在气象大数据云平台数据环境的基础上, 整合各类数据资源, 基于 B/S 架构、SSH 框架, 采用分布式存储技术、数据缓存技术, 建立了跨部门跨行业的气象数据共享平台; 目前, 该平台已经投入业务使用, 为海南省气象数据跨部门跨行业使用提供稳定、高效的实时数据共享服务, 取得了较好的应用效果。

关键词: 气象数据; 数据共享; 分布式存储; 气象大数据云平台; 跨部门

Design and Implementation of Hainan Meteorological Data Sharing Platform Based on Meteorological Big Data Cloud Platform

WANG Shuangshuang, DU Jianhua, WANG Lijun, ZHAO Bing, LIU Jichao

(Meteorological Information Center of Hainan Province, Haikou 570203, China)

Abstract: In order to break the “data island”, promote the integrated use of meteorological data across departments, and give full play to the scientific value of meteorological data, a distribution process flow for the meteorological data cross departments is designed based on the demand type and data format of meteorological data, relying on the special line network, aiming at the problems of messy data sharing mode, low sharing level and lagging management between the meteorological department and other departments in Hainan Province. Based on the data environment of meteorological big data cloud platform, various data resources are integrated. Based on the B/S architecture and the SSH framework, the meteorological data sharing platform with cross departmental and cross industry is established by using the technology of distributed storage and data cache. At present, the platform has been put into operation, which provides the stable and efficient real-time data sharing services for the cross departmental and cross industry use of meteorological data in Hainan Province, and has achieved good application results.

Keywords: meteorological data, data sharing, distributed storage, meteorological big data cloud platform, cross-department

0 引言

随着气象信息化的发展, 我国建成了“天地一体化”的通信系统, 气象数据呈几何级数增长, 各行各业和社会公众对气象服务需求日益旺盛, 涉及旅游、交通、农业、林业、环保、水利、海洋等领域, 气象数据在国家发展和民生服务等领域发挥着日益重要的作用。但长久以来各行业相互独立, 没有形成共享数据的机制, 很难利用大数据技术开展数据挖掘和数据融合, 也就不能更好地发挥数据的价值以及更好服务国计民生。因此, 打破“数据孤岛”, 使气象大数据安全有序地进行流通和共享, 提高数据交互效能, 真正实现数据的融合, 需要进一步研究解决。

近年来, 国内气象部门在气象数据共享方面进行了很

多有益的研究工作, 其中中国气象局于 2009 年启动了全国综合气象信息共享系统建设 (CIMISS, China integrated meteorological information sharing system), CIMISS 系统集成数据收集分发、质量控制、数据解析、存储管理、业务监控为一体, 为国省两级气象部门构建了统一的数据环境^[1], 实现了气象数据的集约化管理。在此基础上, 各级气象部门依托 CIMISS 建设了各部门本地化的共享系统。何林等^[2]以省级集约化气象数据环境及其服务接口为依托, 研发了面向省、市、县三级一线业务人员的统一气象数据共享系统。随着气象数据观测频次不断增高、加工产品的内容和数量不断丰富以及业务系统对数据访问时效的需求更高, CIMISS 系统无法满足对数据进行全面的收集和管

收稿日期: 2022-05-18; 修回日期: 2022-06-17。

基金项目: 国家自然科学基金(41775011); 海南省气象局科技创新项目(HNQXSJ202114)。

作者简介: 王双双(1991-), 女, 河南商丘人, 硕士, 助理工程师, 主要从事气象数据应用方向的研究。

通讯作者: 杜建华(1981-), 男, 湖北洪湖人, 硕士, 高级工程师, 主要从事气象数据理论和应用方向的研究。

引用格式: 王双双, 杜建华, 王立俊, 等. 基于气象大数据云平台的海南气象数据共享平台设计与实现[J]. 计算机测量与控制, 2022, 30(10): 222-226, 232.

理,不能有效满足智慧气象和大数据服务的现实需求。基于云计算、大数据等信息技术^[3-4],2017年国家气象信息中心开始构建气象大数据云平台^[5],支撑国省两级业务^[6],云平台汇聚气象部门内外数据,提供云上加工、挖掘分析的计算服务,开放平台、数据、算法资源,提供数算一体的气象服务,全面支撑“云+端”的气象业务^[7],实现了对CIMISS的全面升级和替代。2020年海南省气象部门搭建了省级气象大数据云平台,支撑海南省一市县两级业务,但更侧重于气象部门内部进行了数据的交换和共享,在行业及外部门的共享探索较少。

随着气象防灾减灾、部门合作、专业专项服务等方面对气象数据的需求日益凸显,海南省气象部门对外提供数据共享服务的范围不断扩大,同时不同的服务对象对共享的数据范围的需求有所不同,如琼港合作项目、三防、部队、环保等。目前的处理方式是各部门提出要求,分别开发实现,气象相关数据对外服务没有统一的平台,造成对外共享数据存在许多系统,共享方式杂乱,不利于对数据使用的管理;前期存在非气象部门直接访问气象部门内网相关资源的现象,不利于气象局内网资源的安全与保密;由于系统分散,对数据分发共享的工作缺乏运行保障,影响气象部门的对外服务形象。所以亟需建立一套气象数据分发平台,实现对外数据共享服务的统一管理、统一出口、统一维护。

本文针对海南省气象部门对外数据共享长期存在的问题,从外部门对气象数据的需求出发,依托气象大数据云平台,基于分布式存储技术和数据缓存技术,实现气象数据共享分发管理的规范化和集约化,文中对气象数据、系统设计、关键技术、系统功能和应用等方面进行介绍。

1 气象数据概述

气象数据资源涵盖气象业务、服务、科研、管理以及行业社会交换的所有数据,包括基础数据、应用支撑数据、管理与运行信息。数据资源种类繁多、格式多样、数据量大。基础数据为气象业务数据资源的全集,包含气象观测数据、气象产品数据、行业社会数据、GIS数据和元数据。气象观测数据是气象直接观测和遥感观测系统所产生的观测数据,包括地面、高空、海洋、辐射、农业和生态气象、大气成分、灾害预警、雷达、卫星、人工影响天气、空间天气、科学考察、历史气候代用等数据,还包括部分数字化数据。气象产品数据是指气象基础统计产品、气象监测数据产品、气象预报业务产品、气象服务产品、气象电子出版物和媒体数据。海南省气象部门与外单位共享数据主要包括地面、高空、海洋、卫星、雷达(基数据)等各类观测数据,以及观测数据生成的基础统计产品如雷达产品(PUP产品、拼图产品等)、卫星产品,以及预报产品、预警信号等。

2 系统设计

2.1 总体架构

海南省气象数据共享平台致力于解决现有为外单位共

享数据方式杂乱、缺乏有效管理的问题,实现共享数据的统一管理,为外单位外部门提供定制化的共享服务。根据海南省气象局气象信息系统集约化管理实施细则(试行),数据环境统一使用气象大数据云平台,不再单独建设数据库,采用现有气象大数据云平台的数据环境,在遵循气象大数据云平台的数据标准和存储规范下,将海南本地特色等资料存入数据库,在其基础上扩建文件索引库表用于存储雷达产品(PUP产品、拼图产品等)、卫星产品索引信息,文件实体保存在分布式NAS和表格存储等。同时建设数据缓存表,用于缓存经常需要调用的数据如整点实况数据等以及卫星云图和雷达产品等较大数据的本地缓存,避免反复读取数据,增加服务器性能的消耗。

系统基于B/S架构^[8-9]和SSH^[10-11]框架,采用JAVA语言实现。系统设计采用四层结构:自下而上依次为基础支撑层、数据存储层、业务逻辑层、界面层。系统总体架构如图1所示。基础支撑层为系统提供所需要的计算、存储、网络资源、安全设备等硬件支撑及相关的软件支撑,依托海南省气象大数据云平台和基础设施资源池实现。数据存储层包括气象大数据云平台业务数据库、本地数据、缓存数据库表、共享目录等,负责存储海南省各类气象观测数据、气象产品、业务管理信息、运行日志等,依托气象大数据云平台和扩展存储环境实现。业务逻辑层以数据存储为基础,包括数据处理、接口管理、系统管理、日志管理、数据监控等。界面层表示系统以web网站的形式实现为外部门提供定制化数据共享服务,可灵活配置满足共享需求,具有可扩展性。

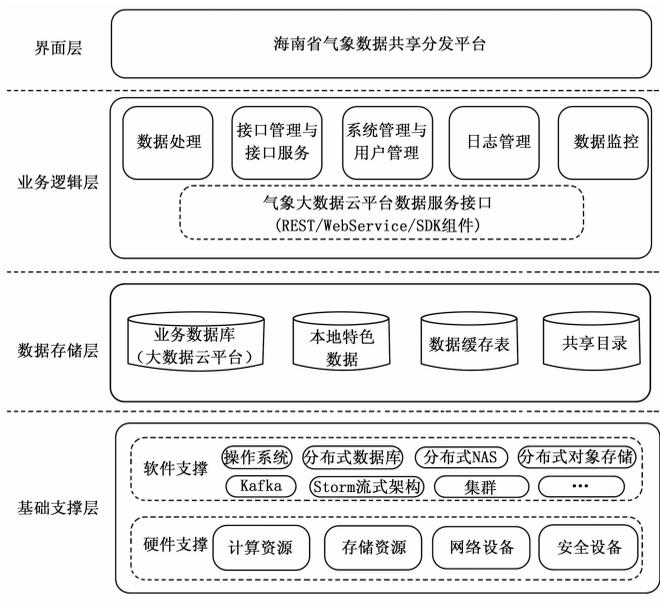


图 1 系统架构

2.2 数据流程

海南省气象数据共享平台依托专线网络、气象大数据云平台数据环境及统一数据服务接口,定时获取气象数据,按照需求推送定制化数据。由于气象数据种类繁多、不标

准、不统一，缺乏和国际、业界一般采用标准格式的一致性，2019 年根据中国气象局统一规划部署，海南省气象部门开始将观测数据由国内自定义数据格式逐步传输、使用标准格式气象数据，目前还未完成所有数据的标准化切换。目前还存在原始格式数据和标准数据同时应用的过渡期，根据需求单位需求数据的不同，文中考虑部分单位还在使用原始格式数据，可以根据需要进行格式转换。根据外单位需求数据类型是原始格式、标准格式还是文件，其中国家站、区域站等地面自动站数据、酸雨、辐射、雷达基数据目前已经传输标准格式数据，可转换为原始格式数据推送。如果后台可通过缓存表读取到所需数据，就从数据缓存表加载并打包相关数据并进行推送，否则就读取相关数据接口或相关共享目录等方式获取并打包推送，同时设置数据监控模块，监控数据是否推送完成。数据流程如图 2 所示。

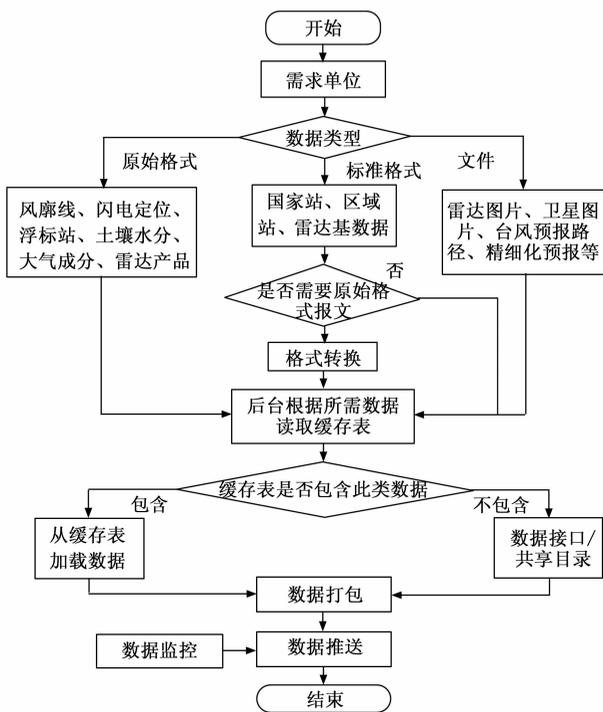


图 2 数据流程

3 功能模块设计

气象数据共享平台主要由 4 个子系统组成，分别是用户管理、系统配置、数据分发配置及数据统计查询，平台功能框架如图 3 所示。

用户管理主要对本地用户进行管理，根据系统定义的不同级别用户的操作权限创建不同的用户，限制用户对系统的操作，以保障用户业务系统的稳定及业务数据的安全。系统配置主要用于对站点管理、用户单位的管理和数据权限的设置，便于系统的扩展。站点管理是将海南省全部地面站、雨量站、交通站、浮标站、船舶站、海岛站、水文站通过站号、站名、所在省市县、是否支持温度湿度降水

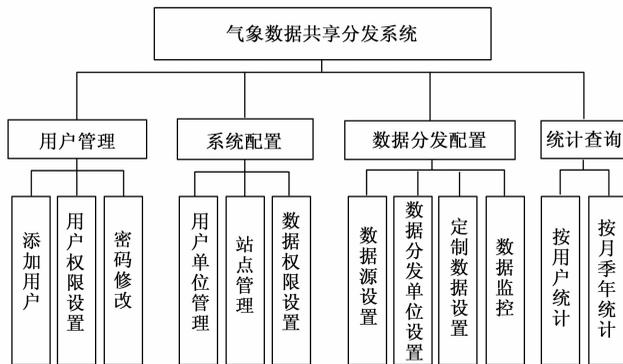


图 3 海南省气象数据共享平台功能模块

风向风速气压能见度六要素，可按照分类及站号定制共享给需求单位；用户单位管理是维护用户单位基本信息，包括外部门负责人员联系方式等；数据权限设置包括数据服务的期限、相关协议（包括保密协议）等，本平台在服务期限范围内向该用户单位提供数据服务。

数据分发设置主要包括数据源设置、数据分发单位设置、定制数据设置、数据监控。目前气象共享数据包括地面、高空、雷达、云图、卫星等各类观测数据，及基础统计产品如雷达拼图产品，以及预报产品、预警信号等。数据源设置是针对不同的数据类型设置获取方式，包括资料名称、分类、获取方式、账户、密码、服务器 IP、接口地址或共享文件目录、时间格式、数据时区、文件类型、更新频率等信息以及数据缓存表存储热点数据或者较大的产品数据的方式，包括设置数据库连接方式、表名等。数据分发单位设置是系统管理员根据管理部门批准的数据分发范围，填写分发用户路径，账户名密码等信息。定制数据设置是按用户单位进行数据范围的授权设置，包括数据的类别（包括产品类别）、自动站站点、所需要素、所需数据频率。为保证系统的稳定运行，以及数据共享分发的时效性，系统通过运行监控模块对数据源的读取情况、数据打包情况以及分发数据是否到达进行监控，实时反馈异常情况，方便运维人员及时处理问题，保障系统稳定运行。

系统根据用户单位所需分发数据的配置，定时自动到对应数据源获取相关的文件，并推送消息给打包服务，按照一定的数据规范进行打包，并推送消息给分发服务，系统自动将需要分发的数据包，推送到指定的目标地址，并检查目标地址数据包是否到达、文件大小是否变化，同时给数据监控模块发送消息。数据统计查询主要用于完成相关的报表工作。

4 关键技术

4.1 气象大数据存储管理技术

气象数据的存储技术与数据形态相关，气象数据可分为结构化数据和非结构化数据，存储技术相对独立。随着大数据时代的到来，数据种类持续增长，数据量激增，传统集中式存储也无法满足业务需求^[12]。随着大数据、云计算技术快速发展，分布式计算日趋成熟。分布式存储技术

具有天生的在线伸缩、高可用、水平扩展等能力,能有效避免单点故障,气象数据以分布式存储技术为主^[13-18]。

对于结构化数据如地面、高空、辐射、农气等站点数据,经过解析、简单处理之后可以用二维表结构逻辑表达,使用关系型数据库表示和存储。分布式关系型数据库、分布式分析型数据库成为结构化数据存储的主流。对于这些站点数据,在分布式关系型数据库中存储近段时间数据(如 1 年)^[19],选用虚谷数据库支撑实时结构化数据的解码入库和存储以及文件索引信息的存储。在分布式分析型数据库中存储长序列的数据,选用列式数据库南大通用 GBase 8a 数据库,对少量的气象要素的检索及分析具有较好的性能。对于雷达、卫星等非结构化数据采用分布式 NAS 存储,支持气象业务中进行大量的统计加工算法直接读写,同时非结构化数据以较大的文件为主,在数据管理和共享访问性能上优势明显。

本系统将海南本地特色等资料按照大数据云平台数据接入流程和标准,根据数据格式规则,将结构化数据解码为要素存入虚谷缓冲库(个别资料直接存入服务库),为了提供数据服务,将数据同步至大数据云平台的虚谷服务库,同时,为了存储长时间序列的数据,还需要将数据从缓冲库同步至 GBase 8a 分析库;根据文件名规则,将雷达产品(雷达拼图产品等)等非结构化数据文件实体保存在分布式 NAS 和表格存储,并按照文件名解析将索引信息存入索引表提供接口服务。对于数据模式预报数据需要把拆分出来的单场单要素的格点数据存储表格存储,重组后的文件存入分布式 NAS 并提取索引存入文件索引库提供接口服务。在索引入库成功和表格存储入库成功之后分别生成对应的 DI 信息,通过本地 DI 容错代理应用程序转发到综合业务监控平台,同时在处理的全过程中有异常或者错误信息生成 EI 告警信息发送综合业务监控平台。

4.2 数据服务接口

气象数据统一服务接口(MUSIC, meteorological unified service interface community),基于大数据云平台数据环境,为国一省一市县各级气象业务和科研用户提供标准的数据访问服务以及应用编程接口,为各类应用系统提供权威的数据接入服务^[20],不仅提供全面的数据获取功能,包括:站点资料的检索、统计,格点资料的解析、裁剪和抽取,文件产品的查询、下载,台站信息的检索等;同时支持本地特色等新增资料接口的发布以及提供数据回写功能,包括站点资料回写、文件产品回写、格点要素回写等。针对气象应用系统提供多种不同的服务方式,包括客户端调用服务、web service、REST 服务和脚本服务,针对不同的服务方式,提供多种调用方法,返回多种可选择的格式,支持内存对象(结构体/类对象)、序列化字符串(XML/JSON/HTML/TEXT 等)、数据文件(XML/JSON/HTML/TEXT 等)、以及 GIS 数据格式(KML)等等,满足不同的应用需求。本系统将海南本地特色等资料存入数据库后,按照气象大数据云平台接口标准实现数据接口配置

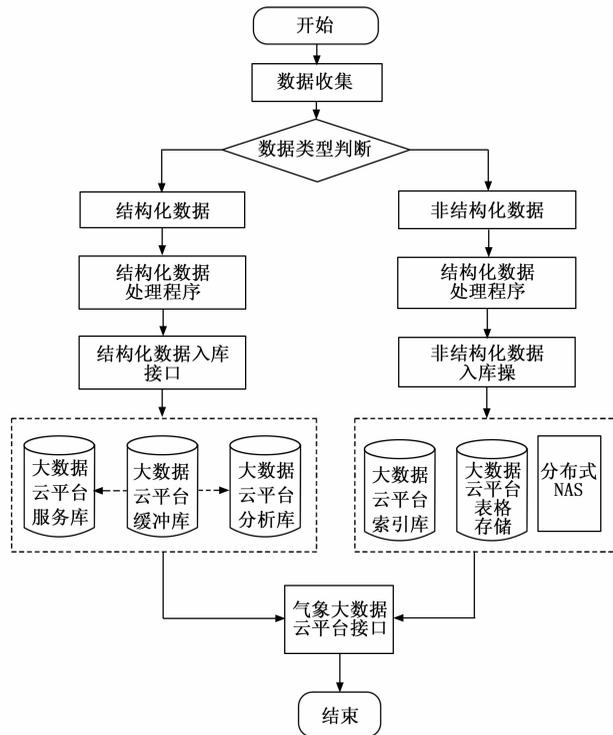


图 4 数据存储流程

和发布,降低业务系统与数据环境间的耦合度,屏蔽底层业务数据库变更带来的影响,确保数据使用规范性和可扩展性。

4.3 缓存技术

气象数据的服务特点有特定的时间、空间规律^[21],呈现冷热不均的分布,比如是整点的地面实况观测数据以及有天气过程的地区数据是需求热点,这样的在需求总量中需求频次较高的数据被称作“热数据”,需求频次较低的数据被称作“冷数据”。海南省气象数据共享分发平台采用缓存技术,将热点数据通过大数据云平台进行一次性调取,存入本地缓存表,同时将较大的雷达产品及云图产品也存入本地缓存表,当给外部门需求单位共享分发数据时优先与速度更快的缓存交互,避免多次调用大数据云平台接口同一种数据,增加省级节点并发访问压力,影响全省业务使用,同时等待返回时间也会增加,命中缓存后有效提高响应时效和需求单位体验。卫星产品及雷达产品缓存表结构如表 1~2 所示。

表 1 卫星产品数据表

字段名称	字段类型	长度	允许为空	默认值	键类型	说明
id	int		否	自增长	pk	序号
satellite_code	varchar	10	否			卫星编码
file_name	varchar	100	否			文件名称
observation_time	datetime		否			资料时间
input_time	datetime		否			入库时间
Satellite_name	varchar	100	否	0		卫星名称
file_size	numeric	10	否			文件大小

表 2 天气雷达产品数据表

字段名称	字段类型	长度	允许为空	默认值	键类型	说明
id	int		否	自增长	pk	序号
radar_code	varchar	10	否			雷达编码
file_name	varchar	100	否			文件名称
observation_time	datetime		否			资料时间
input_time	datetime		否			入库时间
Radar_model	varchar	6	否			雷达型号
file_size	numeric	10	否			文件大小
station_id	varchar	10	否			雷达站号
covering_area	varchar	10	否			覆盖范围
prod_type	varchar	10	否			产品类型

5 系统应用

气象大数据云平台以“云+端”的气象技术体制支撑气象业务系统和各类应用，融数据收集、交换、存储、加工和应用支撑为一体，为业务、服务、政务、科研、培训等提供数据、算法和算力支撑的“数算一体”平台，改变过去业务系统烟囱林立式的建设方式，各业务系统的输入和输出数据均由大数据云平台统一管理，不再独立建设数据库冗余存储，实现硬件、数据、流程、平台、监控的高度集约。气象大数据云平台共享和管理的数据主要包括核心气象数据如地面、高空、辐射、海洋、农气、大气成分、雷达（基数据）、卫星观测数据、气象灾害及预警数据等气象观测数据、加工处理的产品数据如基础观测数据生成的统计产品，雷达产品（PUP 产品、拼图产品等）、多源融合及再分析产品、卫星产品等，以及社会数据、行业数据，如交通、农业、国土、环境、地震、林业、环保、水利、海洋、航空等外部门交换的数据。这些数据汇聚至气象大数据云平台，进行规范的质量控制和加工处理，生产丰富的统计类、格点化、多源融合等产品，进行全生命周期的存储管理，提供标准统一、访问高效的服务接口，对数据和业务的全流程进行可视化监控。气象大数据云平台开放数据交换、产品加工、挖掘分析、数据存储和访问分析等能力，共享数据、算法和接口等资源，支撑全省气象部门的应用和共享服务。气象部门内部业务及科研用户可申请气象大数据云平台存储、算力、接口等资源，生成丰富的产品并提供服务，满足智慧气象和大数据服务的现实需求。

根据上述章节系统设计及关键技术的应用，基于气象大数据云平台的海南省气象数据共享平台构建了集约统一的省级共享气象数据环境，将外部门所需要的海南省本地特色气象资料按照流程和标准接入到气象大数据云平台后，共享各类数据比如地面区域自动站、国家站、海洋浮标、雷达基数据、雷达产品等全部来源于气象大数据云平台，满足气象信息系统集约化要求，实现了为外部门单位提供数据共享分发服务，为海南省气象部门跨部门共享数据提供支撑，满足禁止非气象部门直接访问气象业务内网资源

要求，保证气象部门内部网络可靠安全运行。系统界面展示效果如图 5 所示。



图 5 系统界面展示

6 结束语

针对海南省气象部门与外部门现有数据共享方式杂乱、共享水平低、管理滞后等问题，结合气象数据特点，依托气象大数据云平台，依照气象信息系统集约化要求，设计并实现了海南省气象数据共享分发平台，可为外单位提供定制化的数据服务，为打破“数据孤岛”，推动气象数据跨行业跨部门融合使用奠定了坚实的基础，提高了海南省气象业务内网资源的安全与保密性，规范数据分发共享工作的使用流程，提升了海南省气象部门的对外服务形象。平台目前已投入业务运行，各项功能运行稳定。

参考文献：

- [1] 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.
- [2] 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.
- [3] 冯勇, 李微, 朱辉, 等. 云计算环境下山东省气象大数据云平台的设计与实现 [J]. 信息技术与信息化, 2021 (5): 147-150.
- [4] 卢健, 刘钊. 承载气象业务应用提供数据共享服务气象大数据云平台将投入业务试运行 [N]. 中国气象报, 2020-12-02 (1).
- [5] 张格苗. 气象大数据行动计划出台将建权威在线数据仓库及数据挖掘应用云计算平台 [N]. 中国气象报, 2018-01-03 (1).
- [6] 气象大数据云平台国家一省级业务布局 [J]. 气象科技进展, 2018, 8 (1): 122.
- [7] 刘国强, 熊伟, 汪华. 基于“天擎”的气象核心业务支撑系统优化集约 [J]. 河北农机, 2021: 151-152.
- [8] 杜建华, 王立俊, 王双双. 基于 CIMISS 的海南气象信息服务网设计与实现 [J]. 计算机测量与控制, 2022, 30 (3): 244-250.
- [9] 吴晓珊, 曹旭东, 王森, 等. 基于 B/S 架构的管理系统软件开发 [J]. 计算机测量与控制, 2019, 27 (2): 123-128.
- [10] XINGC F, WANGF Q, SUN Y. A new web application framework based on SSH [J]. Applied Mechanics and Materials, 2015, 3744 (713/714/715): 2203-2207.

(下转第 232 页)