

戚云枫,曾小团,梁苑苑,等. 广西网格预报系统融入“天擎”的实践与思考[J]. 气象研究与应用,2022,43(2):111-116.

Qi Yunfeng,Zeng Xiaotuan,Liang Yuanyuan,et al. Practice and thinking of integrating “Tianqing” into Guangxi grid forecasting system [J]. Journal of Meteorological Research and Application,2022,43(2):111-116.

广西网格预报系统融入“天擎”的实践与思考

戚云枫¹, 曾小团^{1*}, 梁苑苑², 白 龙³

(1.广西壮族自治区气象台, 南宁 530022; 2.广西壮族自治区气象信息中心, 南宁 530022;

3.南宁市气象局, 南宁 530022)

摘要: 分析广西网格预报系统的架构及数据流程、自治区级气象大数据云平台“天擎”规范要求技术特点和差异,提出系统在云化改造各环节中应采用的技术方法,包括数据源整合、数据处理、数据存储以及流程监控等方面关键技术,实现系统完全融入“天擎”。结果表明,改造后数据量提升 3 倍而访问时间减少近 1 倍,性能提升显著。最后,总结了云化改造过程的优点与不足,为后续系统改造和优化提供参考。

关键词: 气象大数据云平台;云化改造;气象信息化

中图分类号: P49

文献标识码: A

doi: 10.19849/j.cnki.CN45-1356/P.2022.2.19

引言

为提升气象信息化事业高质量发展的效益和水平,气象部门近年来构建融合创新、开放共享的气象大数据云平台“天擎”。气象大数据云平台以业务为导向,提供“云+端”的技术架构,可支撑气象业务在大数据中的应用。“云+端”的技术架构在保持应用端多样性的同时,满足数据和算法的统一调配,使得相同业务主线的不同应用系统在应用层面集约。目前气象部门的业务系统大多为集中式或非云端的分布式架构,需要云化改造^[1-4]。

广西智能网格气象预报系统作为天气预报制作核心业务系统,根据精细化格点天气预报业务流程要求设计,全面支撑自治区和市两级网格天气预报的制作和发布。系统按照分层构架方式进行设计和开发,遵循 SOA 架构标准,采用 B/S 架构。系统预设通用流程为自治区气象台下发格点指导预报产品,各市气象台对上级指导产品订正后,实时上传至上级数据中心,拼接处理后形成全区格点预报产品。为体现预报员经验在极端天气预报中的作用,系统

通过服务器集群提供给预报员完成自治区、市两级格点精细化预报交互订正和产品分发功能,实现主客观思路融合。系统还兼顾原有站点和落区预报业务,实现格点到站点和格点到落区的一体化预报业务。系统拥有强大的算法体系,包括预报网格数据处理算法、网格编辑订正算法和气象要素之间的网格一致性处理算法等。

但是,受限于原 CIMISS 数据环境,广西智能网格气象预报系统很多的算法和数据处理放在服务端完成,影响了系统的整体效率^[5]。随着信息化的推进,气象大数据云平台“天擎”为业务实现数据和算法的统一调配提供了可能。遗憾的是,天擎目前仍处于不断建设和完善的阶段,完成云化改造的气象业务系统仍比较少^[6-11],加上各省业务各有特点,尚没见到可供业务实践参考的案例报道。为此,在全面分析广西网格预报系统的架构及数据流程与自治区级气象大数据云平台“天擎”规范要求技术特点和差异后,发现广西智能网格气象预报系统在一些标准上与“云+端”的气象技术体制类似,且为全面最具代表性的核心业务系统,优先开展该系统云化改造对

收稿日期: 2022-05-10

基金项目: 广西自然科学基金项目(2022GXNSFAA035482)、广西气象科研计划项目(桂气科 2022ZL04)

作者简介: 戚云枫(1990—),男,工程师,主要从事天气预报技术研究和开发工作。E-mail: fanie-yani@qq.com

通讯作者: 曾小团(1971—),男,正研级高级工程师,主要从事天气预报技术研究与预报系统开发工作。E-mail: 158083890@qq.com

推进全区业务系统的全面集约升级可起到示范引领作用。本文对系统改造实践进行全面总结,为相关业务提供技术参考。

1 系统云化改造思路

系统云化改造需遵循气象大数据云平台“天擎”的思路设计,按照“云+端”的标准重新设计系统的技术架构。在“天擎”中,“云”的部分通过基础硬件、虚拟化资源、软件算法以及网络组成,满足硬件集约的要求,同时可为融入系统的高效、稳定、安全运行作重要保障,而“端”的部分注重用户交互,提供服务。原智能网格气象预报系统在4台操作系统为Windows Server 2012的服务器上部署,软件环境包括JDK、Nginx、Tomcat等中间件,由数据同化、数据加工处理、应用支撑和应用服务等4个部分组成系统。系统架构调整后,基础设施部分由“天擎”支撑,数据与逻辑部分依托“天擎”实现,是云化改造的发力点,端的应用部分保留,但需要进行前端应用流程改造。如图1所示,实线框部分为“天擎”提供的功能,虚线框部分为基于“天擎”研发的内容。

从“天擎”的设计角度出发,一是“天擎”支持且系统需要的模块按照“天擎”的规范设计融入,如软硬件资源、数据环境和算法等,二是“天擎”要求有但系统还没有的模块需要根据“天擎”的标准建立和完

善,如产品监控等。基于“天擎”的设计思路,本系统融入“天擎”的内容主要包括数据及存储的设计、数据处理融入、端的改造和监控融入共四个方面。

数据及存储的设计包括数据源的接入和数据存储设计。数据源接入是将系统的所有数据的获取方式统一改造为从“天擎”读取。本系统用到的数据源包括作为参考资料浏览的地面观测资料、卫星产品、雷达产品,以及作为网格预报背景场的中央气象台网格预报指导产品、数值模式产品和客观预报方法产品等,这些数据由从CIMISS或MICAPS分布式数据环境读取调整改造为从“天擎”获取。数据源接入后,经过数据加工和处理,系统会产生中间数据和产品数据,“天擎”已经为这些数据设计了各种类型的存储。本系统存储的数据包括中央气象台指导产品、数值模式产品、客观预报方法产品和主观预报产品等预报网格,按照业务规范要求的最终网格预报产品、报文产品和图形产品等文件,以及系统使用的用户部门信息、预报员信息、初始场缺省方案、产品索引和站点预报配置等信息。这些数据在原系统中使用关系型MySQL数据库存储,根据“天擎”对数据存储的要求,网格预报数据采用“天擎”实时应用库存储,预报文件采用“天擎”NAS文件系统存储,而系统配置信息采用“天擎”云数据库存储。

数据处理融入是指系统内部的固定算法提交至“天擎”的算法库,通过“天擎”加工流水线进行统一调度管理^[12]。需要理清系统在运行过程中使用的所有算法,明确算法的输入、输出以及流程。对系统运行流程拆解,将预报产品同化处理、衍生要素计算、初始场产品加工和预报产品生成等流程凝练成算法,通过“天擎”的加工流水线统一调度和加工处理。预报产品同化处理算法是将数值模式产品、中央气象台指导预报以及客观方法等预报背景场统一插值成业务规范要求的时空分辨率网格,并存储到“天擎”实时应用库。衍生要素计算是通过原始预报背景场进行日降水量、日最高最低气温和日极大风等统计,初始场产品加工是根据智能网格预报初始场配置方案进行初始化,形成网格预报的初始场。产品生成算法是将预报员制作的预报网格根据业务要求生成网格预报产品文件以及城镇预报、乡镇预报、海洋预报和面雨量预报等报文产品。

“端”的改造是指对用户交互界面的流程

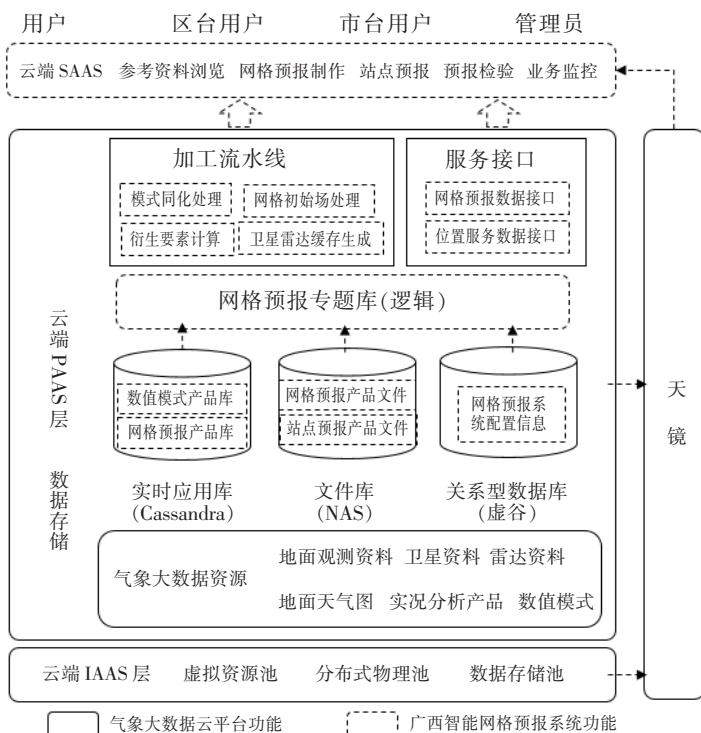


图1 融入“天擎”后的系统架构

改造,直接反映业务系统的应用功能。本系统将数据和算法抽离并放入“云”,“端”的部分剩下 WEB 服务,包括参考资料浏览、网格编辑订正与发布、站点预报生成与发布、系统配置和数据接口等用户交互模块。通过应用服务框架对接大数据云平台,建立 WEB 服务到“天擎”的消息通信和数据交换,提供数据共享接口实现网格预报产品的访问,达成“端”的改造。

监控融入是指系统各个关键环节都需要监控并加入到气象大数据云平台的监控系统——“天镜”。系统改造前仅有产品生成的监控,需将各个关键的环节加入“天镜”的监控。系统监控模块,根据已经完成的系统流程定义监控信息,基于“天镜”开放的监控信息接口,定制个性化的监控界面,达到对网格预报系统硬件资源、数据、产品和流程的全业务、全流程监控。

2 系统云化改造关键技术

2.1 数据源整合

数据源整合解决数据来源的问题。云化改造前,

系统的数据来源有 CIMISS、MICAPS4 分布式数据环境和文件共享三种。

对于 CIMISS 接口的数据来源,申请“天擎”对应的接口访问权限,可以直接读取相应数据。地面观测资料、天气图和物理量、雷达产品和卫星产品等资料可以直接读取“天擎”接口并做前端显示。

对于来源 MICAPS4 分布式数据环境的预报数据,如 CMA-GFS、CMA-GD、CMA-SH、ECMWF 等 13 类模式产品,更新数据源为气象大数据云平台数值预报模式基础数据库或 MICAPS4 分布式数据库。

对于从文件共享渠道获取的客观预报产品,如中央气象台指导预报 SCMOS、南宁 MOS 气温产品、最优滑动气温产品、频率订正降水产品和多模式降水产品等 8 类客观产品,调整为直接从“天擎”NAS 读取。由于客观方法产品多源异构,通过数据湖的方式存入 NAS,再进入加工流水线。如图 2 所示,为系统数据流程图。

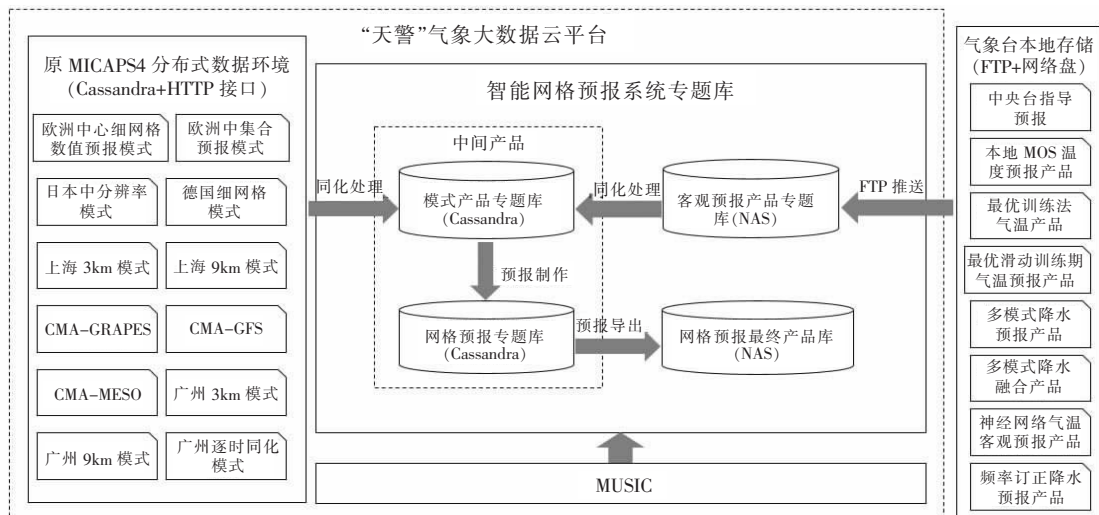


图 2 系统数据流程

2.2 数据处理流程

改造前本系统的数据加工流程主要涉及数值模式同化处理、模式衍生要素计算、网格预报初始场处理和网格预报产品制作等,采用 Windows 和 Quartz 的定时任务框架两种方式进行调度。两种调度方式都有弊端,Windows 定时任务稳定但无可视化页面给用户,Quartz 虽然可以通过开发提供可视化交互页面给用户但可靠性不够高。经过改造将完全使用“天擎”的加工流水线取代现有的任务调度方式。

具体的加工流程改造包括三个步骤:算法抽象、环节设置和任务定义。

算法抽象是将系统的数据加工算法剥离成一套可执行程序。首先,对算法进行梳理、整合和封装,开放程序参数,形成可独立运行的执行程序;然后,确定程序的输入输出。输入与输出需要对接“天擎”的存储环境,尽量采用直连数据库方式访问;接着,向程序添加 DI、EI 监控信息的发送。主要的 DI 信息包括算法执行时输入数据获取、算法计算过程信息和算法输出产品保存等信息,主要的 EI 信息包括数据

获取异常,算法计算过程异常和算法输出保存异常等;最后,将程序注册到天擎的产品加工流水线中。

环节设置是对算法的参数、触发方式等信息进行配置,每个环节承担一整套该算法执行的流程。例如数值模式同化算法,通过参数输入判断对多种的数值模式和客观预报产品的不同要素的处理。环节的触发方式可设定为定时触发、顺序触发和数据源感知触发。

多个环节组成一个任务。在“天擎”的任务管理界面,除了自动执行算法环节外,还可对所有环节的运行手动调度,统一管理。

本系统最主要的数据是半结构化的数值模式和网格预报产品,实现的数据处理算法有数值模式同化算法、客观方法同化算法、衍生要素融合计算和产品生成算法,均使用定时触发方式调度。如图3所示为系统产品加工算法流程图:

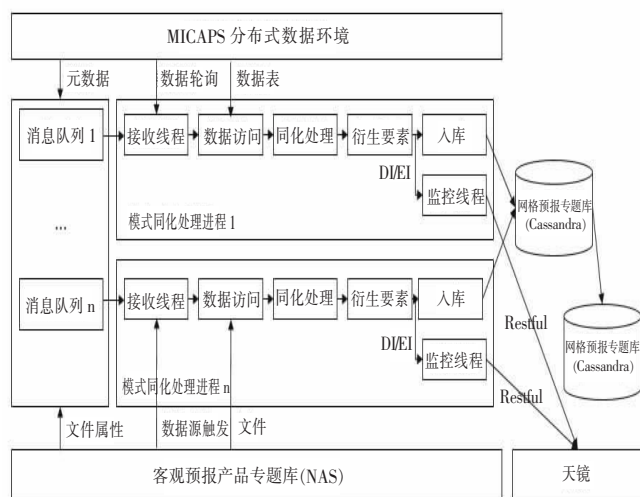


图3 系统产品加工算法流程

2.3 数据存储管理

数据存储管理的改造是将本系统所有由本地MySQL数据库存储和管理的数据,改造为基于气象大数据云平台存储管理。如图4为“天擎”数据存储方案,根据方案设计,系统需构建不同的存储,包括数值模式产品专题库、智能网格预报专题库、基于云数据库(虚谷)建设系统应用配置库、基于云数据库(NAS)建设客观预报产品和最终网格产品库。

网格预报数据作为核心业务数据存放于有6个节点的Cassandra分布式实时数据库,该数据库能满足多用户高并发访问。分布式实时数据库基于键(Key)可快速获取数据块,非常适用海量非结构化

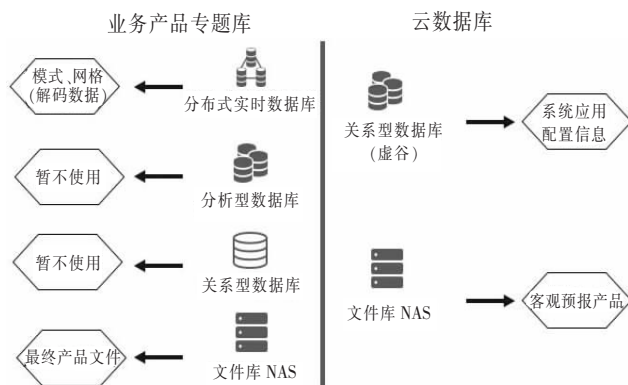


图4 “天擎”数据存储方案

数据存储^[13]。数据库采用按预报名和预报要素集中存储,即一种预报或要素一张表,实现对空间位置和格距固定不变,属性变化频繁获取的四维、五维时态网格数据集模型。如表1所示,该数据模型结构简单,具有按时间/时效存储管理和快速检索优势,适用于数值模式产品和网格预报产品。

表1 网格数据Cassandra表结构

Cassandra 术语	描述	数据类型	示例
表名	模式名称	字符串	ECTHIN
分区键	要素名称	字符串	TEM
集群键	起报时间	日期时间	2021-07-20 08:00:00
集群键	预报时效	整型	3
集群键	层次	整型	999
普通列	压缩后的网格值	二进制数据	010101010101010101

使用虚谷云数据库建立应用配置库,管理系统的配置信息,保存网格预报业务各功能模块所使用的配置。相关算法和后台服务,也可以通过直连数据库的方式,获取数据库中的元数据信息。

使用NAS文件云库分配共享文件存储,存储最终输出的文件产品。在“天擎”的分布式NAS存储系统中根据网格预报业务需求,单独划分一部分存储空间作为网格预报和站点预报历史产品、图形产品的文件库,其采用的存储标准和管理方式,由系统根据网格预报业务规范来决定。

2.4 前端应用流程

配合大数据云平台后端的融入,系统前端应用基于“天擎”的改造主要包括以下两个方面:一是重写后台数据访问和数据管理接口。由于数值模式产

品和网格预报产品存储调整为“天擎”实时应用库,系统配置信息存储调整为“天擎”云数据库,相应的数据接口需要调整。二是建立网格预报缓存机制。为了减少数据库 I/O,提高系统运行性能,采用“天擎”云数据库(Redis 集群)构建网格预报产品数据内存级缓存机制。预报员在网格预报制作过程中,所有编辑订正算法计算的网格数据均位于 Redis 内存,可大大提高网格预报性能。当预报制作完成后,从 Redis 一次性回写 Cassandra 数据库。

2.5 基于“天镜”的监控新流程

“天镜”监控,包括监控信息的发送和读取两个部分。监控信息可以利用“天镜”的 Agent 收集服务器集群的 CPU、内存、磁盘、网络等基础信息,或者通过加工流水线和后台服务推送 DI/EI 日志及自定义监控信息。监控信息可集成在“天镜”众创快速接入框架统一展示,或自行读取其进行可视化展示。

本系统新增业务监控模块,对系统资源、各个业务环节的监控形成日志,以标准接口的形式组装 DI 信息,推送至“天镜”系统,通过自主研发界面显示监控信息。监控内容包含用户行为信息、数据加工处理流程日志、产品制作过程日志、数据共享日志、数据环境情况等。因此自定义的 DI/EI 信息有预报初始场同化处理 DI、用户访问 DI、用户提交网格预报 DI、用户提交站点预报 DI、产品导出与上传 DI 和集中告警 EI。

3 系统改造前后性能对比

改造前原系统为网格预报业务规范的 5km 分辨率网格,改造后将网格提高为 2.5km 分辨率,模拟用户下载网格预报数据得到表 2 中的数据对比。从表格可以看出,改造后单个网格场的数据大小比改造前多了 3 倍,但传输时间快了将近 1 倍。

表 2 系统改造前和改造后网格下载对比

指标	改造前的 5km 网格场	改造后的 2.5km 网格场
经向格点数	161	321
纬向格点数	141	281
总格点数	22701	90201
数据大小	176kb	705kb
查询总耗时	100~250ms	50~150ms

4 总结与思考

广西智能网格气象预报系统按照气象大数据云平台的架构进行云化改造,实现了数据源、数据处理、数据存储和系统监控等全面融入,在预报业务主线一定范围内达到集约。系统已经进入业务试运行阶段。但在建设和业务试运行过程中仍存在许多值得思考和改进的地方:

理、数据加工流水线和流程监控等功能,但“天擎”现阶段的版本缺少众创接口的应用,在处理用户交互的算法中比较困难。同时,“天擎”的中间件资源的稳定性难以掌握,如 Redis 在使用的过程中曾出现过异常,但在“天擎”的管理界面难以看出其异常情况,需要多方排查才能发现问题所在。

最后,气象业务技术体制改革的组织管理处于起步阶段。气象大数据云平台是气象信息化的一个标志,在推进业务系统云化改造的过程中所参考的流程和规范不够全面,同时“天擎”的建设对气象业务系统研发的技术门槛提升了一个档次,这对气象业务人才培养提出了新的要求。将气象业务系统大范围、全面融入“天擎”是一个非常艰巨的任务,需要各方形成合力,做好前期准备。

参考文献:

[1] 任晓炜,梁苑苑,陈婧霆,等.广西气象信息网络业务发展与展望[J].气象研究与应用,2020.

首先,基于气象大数据云平台支撑的业务主线在顶层设计不够成熟。广西智能网格预报业务系统目前仅支撑了短期预报制作、发布和服务等功能,所有的数据、算法、监控和交互等是围绕此系统功能研发,并没有以整个预报业务主线为出发点设计。以该系统的云化改造为契机,验证或者前期探索了预报业务整体上云。

其次,气象大数据云平台的功能的实用性和运行的稳定性未能做进一步的测试。“天擎”提供了一套完善的集约化系统平台,有完善的数据存储和管

- [2] 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):4.
- [3] 李有华,卢小凤,陈剑飞,等.基于智能网格产品的广西行业气象服务集约化系统的设计与实现[J].气象研究与应用,2019,40(4):4.
- [4] 詹利群,任晓炜,黄志,等.广西气象业务内网功能设计与实现[J].气象研究与应用,2019,40(1):6.
- [5] 曾行吉,任晓炜,宋瑶,等.微服务在气象数据服务中的应用研究[J].气象研究与应用,2019,40(1):4.
- [6] 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.
- [7] 唐卫,渠寒花,霍庆.交管风险管控平台与气象大数据云平台融合的实践与思考[J].气象科技进展,2020,10(5):34-41.
- [8] 赵冰燕,郭彩莲,来志云.基于青海气象大数据云平台的数据服务接口[J].青海科技,2021,28(1):82-86,90.
- [9] 雷鸣.气象大数据分布式存储设计与实现[J].计算机技术与发展,2021,31(5):193-197.
- [10] 刘媛媛,何文春,王妍,等.气象大数据云平台归档系统设计及实现[J].气象科技,2021,49(5):697-706.
- [11] 秦运龙,王迎迎,张冰松,等.省级外网气象大数据服务平台研究与实现[J].气象科技,2020,48(6):823-828,854.
- [12] 黄志,黄珩,梁维亮,等.基于“天擎”DPL 的业务融入设计与应用初探[J].气象研究与应用,2022,43(1):73-77.
- [13] 杨永毅,赵思亮,赵芳.气象大数据在 Cassandra 数据库的应用研究[J].现代科学仪器,2020(3).

Practice and thinking of integrating “Tianqing” into Guangxi grid forecasting system

Qi Yunfeng¹, Zeng Xiaotuan^{1*}, Liang Yuanyuan², Bai Long³

(1. Guangxi Meteorological Observatory, Nanning 530022, China;

2. Guangxi Meteorological Information Center, Nanning 530022, China;

3. Nanning Meteorological Bureau, Nanning 530022, China)

Abstract: This paper analyzes the architecture and data flow of the Guangxi grid forecasting system, and the technical characteristics and differences of the "Tianqing" specification requirements for the autonomous region-level meteorological big data cloud platform. The technical methods that the system should adopt in each link of cloud transformation are proposed, including key technologies in data source integration, data processing, data storage, and process monitoring, so as to realize the complete integration of the system into "Tianqing". The results show that after the transformation, the data volume is increased by 3 times and the access time is reduced by nearly 1 times, and the performance is improved significantly. Finally, the advantages and disadvantages of the cloud transformation process are summarized, which can provide reference for subsequent system transformation and optimization.

Key words: meteorological big data cloud platform; cloud transformation; meteorological informatization