

霍庆, 何文春, 何林, 等. 气象大数据云平台算法集约化环境设计与应用. 应用气象学报, 2024, 35(1): 80-89.
DOI: 10.11898/1001-7313.20240107

气象大数据云平台算法集约化环境设计与应用

霍庆¹⁾ 何文春^{1)*} 何林²⁾ 高峰¹⁾ 陈士旺¹⁾ 徐拥军¹⁾

¹⁾(国家气象信息中心, 北京 100081)

²⁾(陕西省气象信息中心, 西安 710014)

摘要

气象业务系统集约化发展和“云+端”业务技术体制改革是实现气象业务高质量发展的重要措施。2020 年中国气象局提出构建以气象大数据云平台为云、气象业务系统为端的“云+端”业务技术体制,明确气象大数据云平台作为关键基础技术平台的定位。加工流水线作为气象算法的集约化环境,应用数算一体、高效任务调度、可视化流程编排、容器等技术,实现气象算法的统一管理与高效集约调度运行。2021 年加工流水线业务运行,支撑全国 202 个业务系统的实时运行,业务系统性能提升 1~10 倍,集约化程度显著提高,对提升业务系统的运行效率、增强业务系统的协同性、加速“云+端”业务技术体制改革进程和推进气象业务集约发展发挥了重要支撑作用。

关键词: 加工流水线; 算法库; 任务调度; 数算一体; 集约化; “云+端”业务技术体制

引言

气象信息化是气象现代化的重要标志^[1]。经过多年发展,我国气象部门气象数据传输、处理、加工、存储与服务等业务能力不断增强,气象信息化水平不断提高。国家气象信息中心先后建立了国家级气象资料存储检索系统(MDSS)和全国综合气象信息共享系统(CIMISS),逐步实现气象数据环境在全国的统一^[2-5],具备一定的数据产品加工能力^[6],一定程度上实现数据的统一供给。随着气象业务的发展,各业务系统纷纷构建自己的产品加工系统和配套数据管理系统^[7-16],但也产生了一个难以回避的问题,即数据孤岛与应用烟囱,是气象业务不集约的表现^[17-18]。

气象业务的不集约会带来诸多问题^[19-22]: ①针对相似的应用场景,数据标准不统一,运维难度大。②各业务系统自建数据库,数据重复存储,造成巨大浪费。③数据本地化存储,因同步不及时导致数据

不一致。④业务系统独立开发与运行,导致严重的信息与技术壁垒,上下游业务无法联动。

近年中国气象局陆续印发多个气象信息化顶层设计,其中《气象大数据行动计划(2017—2020 年)》^[23]首次提出建设气象大数据云平台,明确以气象大数据云平台为云、气象业务系统为端的“云+端”业务技术体制^[24-26],实现数据管理、加工处理、应用服务的高度集约,促进业务系统有效衔接、高效协同。国家气象信息中心于 2018 年设计建设气象大数据云平台(简称天擎),2021 年 12 月实现业务运行。天擎设计定位为全国气象业务的基础性平台,基于混合云提供数算一体的平台化服务,系统包括交换及质量控制、存储与服务、加工流水线、门户网站等^[27]。

加工流水线是天擎实现算法统一管理与集约运行、消除应用烟囱、实现“云+端”业务技术体制的重要基础。本文在分析业务需求的基础上,详细介绍加工流水线的功能设计、关键技术、算法云化集成、应用效益,并对未来发展方向进行展望。

2023-11-20 收到, 2023-12-12 收到再改稿。

资助项目: 国家发展改革委工程建设项目(发改农经[2019]1987 号), 国铁集团科研计划课题(N2022T011)

* 通信作者, 邮箱: hewc@cma.gov.cn

1 业务需求

气象业务对数据产品加工的需求,可以分为 3 个层面:①单一业务系统更关注产品生成效率和稳定性,要求算法高效调度、稳定运行,及时生成数据产品。②不同业务系统之间要求数据产品共享,实现协同联动。③全局气象业务期望集约高效,形成“云+端”业务技术体制,提高整体运行效能。

1.1 数据产品加工高效及稳定需求

气象数据产品及其加工算法是气象业务的核心,算法的高效、稳定调度运行是保障产品能否及时、稳定生成的基础条件。算法调度需要建立集约、高效的调度框架,实现定时、实时等多种调度策略。数算一体是算法高效运行的重要手段,算法靠近数据,减少数据搬运,提升数据的读写效率及算法运行时效。

1.2 业务系统数据共享及协同需求

数据、算法、算力是气象业务的基础三要素,三者协同共享对提升气象业务效率和实现高质量发展至关重要。应用烟囱和数据孤岛导致数据产品仅在单一业务系统内循环,有待互联互通,实现不同业务系统间的数据共享;数据共享流程亟待优化,减少传输、下载等流转环节,实现数据生产者的直存与数据需求者的直取。共享先进算法可以减少重复研发,缩短业务系统的上线周期,统一管理算法,支持用户

获取较全面的气象算法清单。在保障各业务系统资源隔离的基础上,实现算力资源充分共享是提升业务效能的重要手段。

1.3 业务应用高效集约运行需求

构建以天擎为云、气象业务系统为端的“云+端”业务技术体制,是实现气象业务集约化和高质量发展的必由之路。集约的数据环境和算法环境是云的核心能力,算法集约需实现气象算法统一管理,支持版权控制下的共建共享,同时需具备丰富的算力资源和高并发的任务调度能力,支撑算法的高效调度运行。

2 技术设计

2.1 总体架构

加工流水线采用面向服务架构(SOA)的设计思想^[28],各功能模块实现面向服务的封装,系统架构如图 1 所示。依托于天擎的算力资源,加工流水线以算法库和任务管控为核心,与天擎数据存储共同构建数算一体的算法环境。算法库实现算法的标准化,统一管理算法并共享;任务管控支持多种调度策略,以高可靠性与容错性,支撑算法的高效、稳定调度运行。加工流水线面向应用前端将所有功能以接口形式开放调用。基于气象业务综合监控系统(简称天镜),实现算法运行状态的自动采集与异常告警。

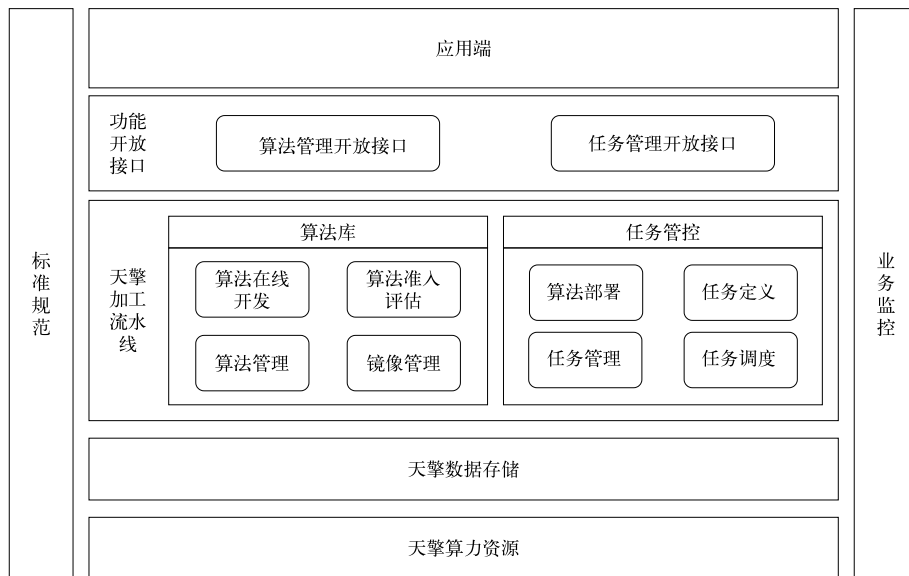


图 1 加工流水线框架

Fig. 1 Framework of data processing line

2.2 功能设计

2.2.1 算法管理

算法管理实现算法注册、算法审核、算法版本管理、算法清单与算法检索、算法版权控制等,支撑用户对气象算法的发现、查看与共享,加工流水线算法管理流程见图 2。算法注册实现算法向算法库的提

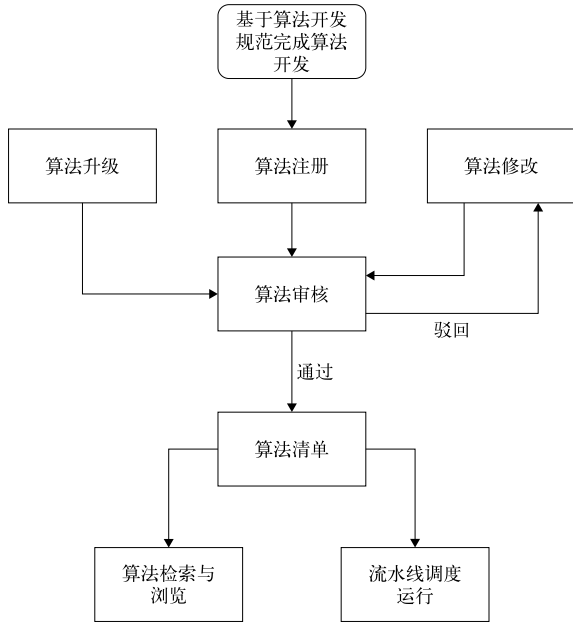


图 2 加工流水线算法管理
Fig. 2 Algorithm management of data processing line

交,同时填报算法元数据。算法通过审核后进入算法清单,供用户检索查看。算法管理实现严格的版权控制,用户向算法所有者申请使用权后才能使用。

2.2.2 任务定义

任务定义主要实现算法运行参数、算法运行节点、算法调度策略、算法日志管理策略、算法监控策略等配置(图 3)。通过新增任务确定任务名称,向任务中添加算法并指定运行节点,实现任务、算法和算力资源的关联。当一个任务由多个算法组成时,可以配置算法的上下游关系,确定算法运行时序。任务定义的核心是完成算法调度策略配置,选定调度方式并定义算法启动时间等。算法监控策略配置包括算法调度运行监控策略和产品完整性监控策略配置,实现算法运行与天镜的关联与监控告警。任务定义支持用户以可视化页面操作的方式进行。

2.2.3 任务调度

任务调度根据任务定义中配置的调度策略实现算法的调度运行。任务调度是加工流水线的核心功能,支持多种任务调度策略,包括定时调度、人工调度、实时调度、顺序调度和数据源感知调度等。其中,数据源感知调度实时分析数据到报状态,在数据达到的第一时间启动算法。

任务调度由调度服务端、调度客户端(即计算节点)和消息中间件组成。调度服务端基于算法信息

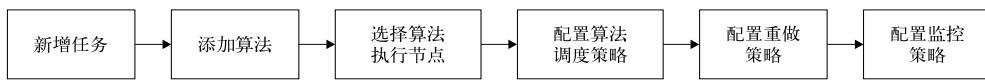


图 3 加工流水线任务定义
Fig. 3 Job definition of data processing line

(包括算法 ID、算法参数、算法执行命令、算法执行所需 CPU 和内存等)、任务信息(包括调度策略、任务可执行的计算节点等)和计算节点负载状态生成调度指令,下发至调度客户端,调度客户端基于调度指令启动并执行算法,将算法执行状态、计算节点状态等通过消息中间件反馈至调度服务端。任务调度流程见图 4。

2.2.4 任务管理

任务管理实现任务启停控制、任务运行状态和日志查看以及任务人工补做。任务运行状态和日志查看页面展示任务的计划执行开始时间、实际执行开始时间、执行结束时间、执行计算节点以及执行日志等。

2.2.5 功能开放接口

功能开放接口将加工流水线的主要功能以接口形式发布,支持应用系统前端直接调用,可以更有效地支撑“云+端”业务技术体制的建立,如图 5 所示。加工流水线的功能开放接口包括算法管理接口和任务管理接口。

2.2.6 可靠性设计

加工流水线的安全稳定运行对气象业务至关重要。从技术层面采用 3 条措施提高加工流水线的安全稳定性:①调度服务端采用主备双节点设计,实时监听进程运行状态和调度服务可用性,发生异常时主备自动切换。②计算节点基于容器技术,当某个计算节点出现故障时,实现算法向其他节点自动迁

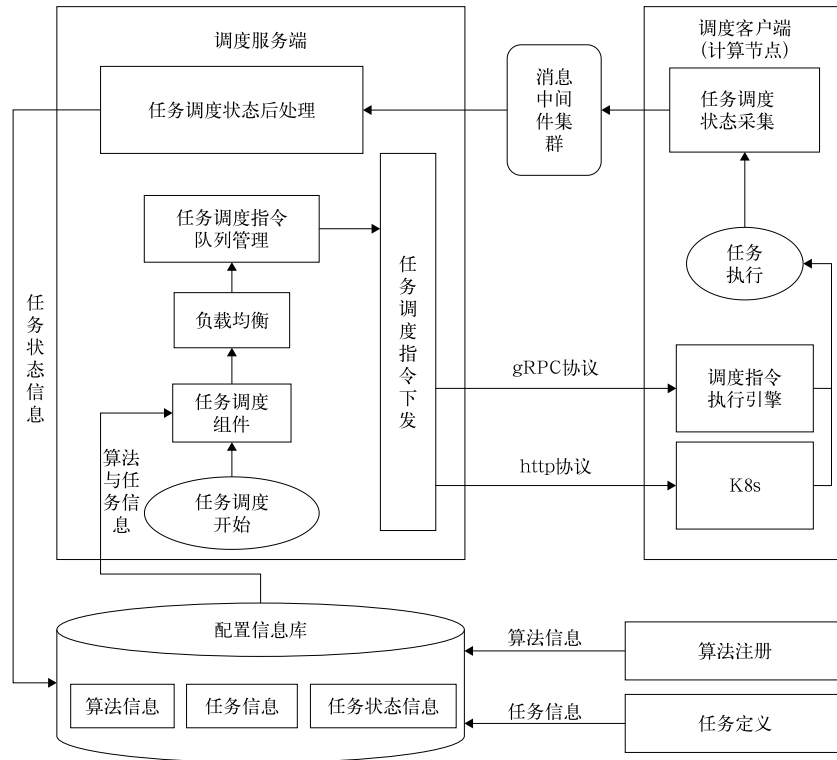


图 4 加工流水线任务调度

Fig. 4 Task scheduling of data processing line

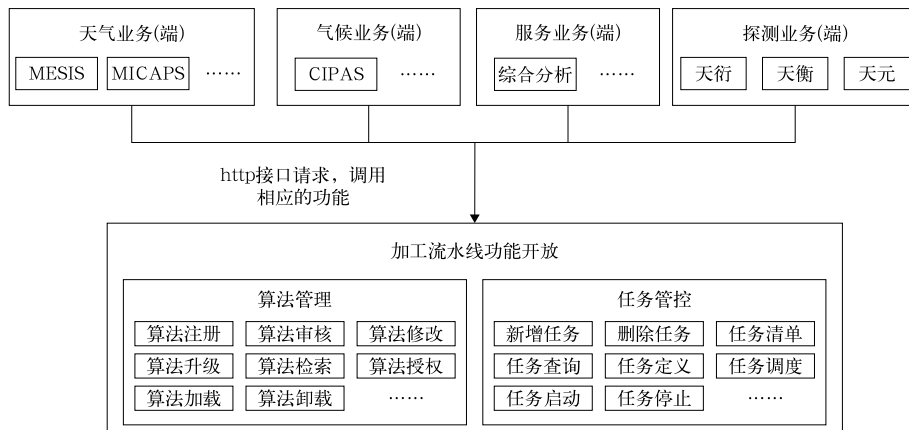


图 5 基于加工流水线功能开放接口构建“云+端”业务架构

Fig. 5 “cloud+end” business architecture based on open functional interface of data processing line

移。③基于天擎备份环境,算法在主备环境同时运行,异常状态下应用系统可实现主备切换。

3 关键技术

3.1 数算一体技术

加工流水线实现气象数据资源、算法资源、算力

资源的一体化,数算一体的特性从 3 个方面予以保障:①算法靠近数据,天擎同时承载气象数据管理与气象算法运行,算法向加工流水线部署的过程,即算法向数据靠拢的过程,减少数据搬运与传输。②加工流水线上的算法可以访问数据库,不用通过调用服务接口即可直接读取或写入数据。③存储环境与加工流水线的计算环境在物理部署上高速互联。加

工流水线的数算一体特性显著提升气象数据产品的加工时效。

3.2 容器技术

天擎投入运行与应用前,气象业务系统主要以虚拟机或物理机方式运行。物理机无法实现资源隔离,不同业务间会出现资源抢占,影响业务运行稳定性。虚拟机运算性能相对较低,一般用于支撑系统前端^[29]。相比于物理机,容器实现了资源隔离,业务系统部署与迁移更便捷;相比于虚拟机,容器具备运算性能高、资源利用率高、启动耗时短等特

性^[30-37]。

加工流水线的容器集群管理基于 Kubernetes (K8s)实现,将集群系统分为管理节点和计算节点^[38-39](图 6)。管理节点接收加工流水线调度服务端的调度指令,基于 API 管理服务、容器控制服务、容器调度服务、数据存储组件等实现集群的资源管理、容器启停与扩缩容、系统监控等。计算节点基于容器执行服务、容器网络管理等组件,完成容器创建与容器状态采集。管理节点和计算节点通过超文本传输协议 http 进行通信。

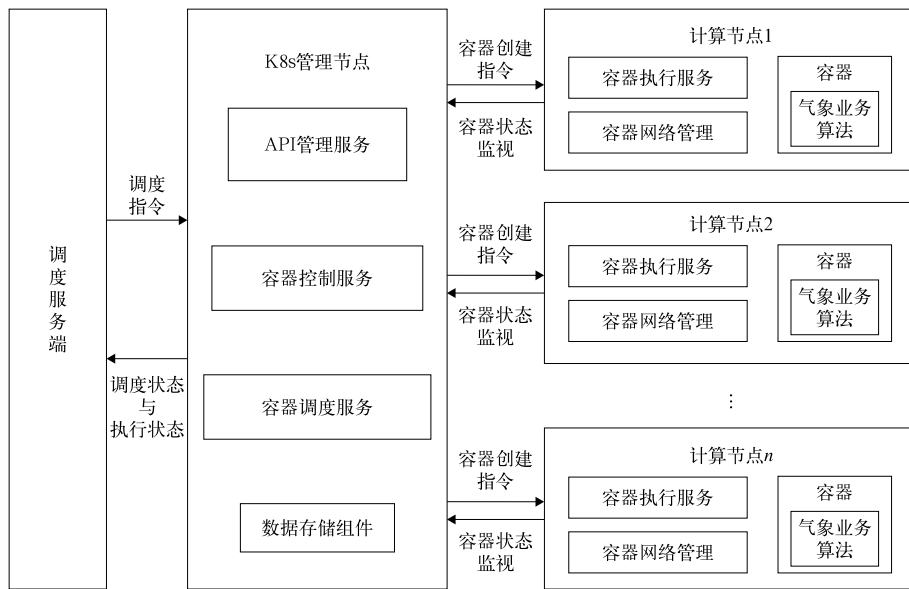


图 6 加工流水线容器技术

Fig. 6 Key technology of data processing line-docker and K8s

3.3 加工流程可视化技术

气象数据产品加工任务一般由多个算法串接而成,如天气雷达产品加工任务包含单站质量控制产品生成算法、单站图片产品生成算法、组网产品生成

算法和组网图片产品生成算法(图 7)。加工流水线基于 mxGraph 实现可视化任务流程的创建、修改、编辑、串接等^[40]。

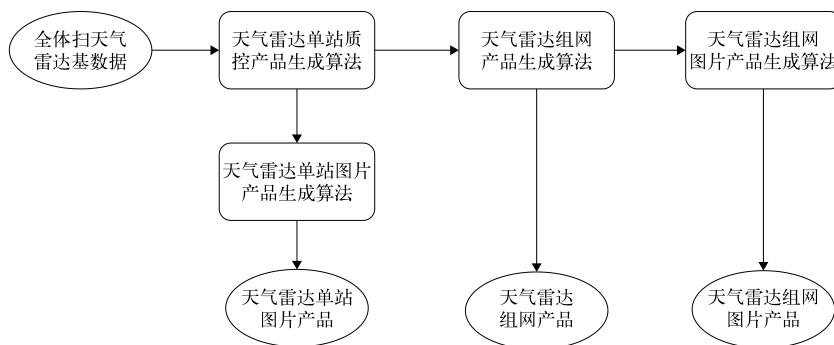


图 7 加工流水线天气雷达产品加工任务流程示意图

Fig. 7 Task flow of weather radar products on data processing line

4 算法云化集成

按照“云+端”业务技术体制的要求,前端应用部署在虚拟资源中,后端产品加工算法按照规范进行云化改造并部署于天擎上运行。算法云化要求输入数据来自天擎、输出产品回写至天擎且进行产品注册发布、算法由加工流水线统一调度运行、天镜统一监视运行等。

算法云化集成的核心流程包括:天擎业务账户与数据访问权限申请、算力资源申请、数据产品注册、算法开发测试、算法镜像封装和算法集成运行。账户与数据访问权限申请、算力资源申请和数据产品注册等参见《天擎应用开发手册》,本文主要介绍算法开发测试、算法镜像封装、算法集成运行的技术操作。

4.1 算法开发测试

基于天擎临时容器资源进行算法开发测试,且须符合一定的开发规范,包括应用加工流水线的公共配置文件、输入数据与输出产品对接天擎存储、监视信息对接天镜等。天擎主备环境及省级节点各自的公共配置文件存储着该节点的数据库 IP、服务接口域名、天镜监视域名等公共信息。应用公共配置文件使算法在不同节点间的迁移部署更便捷,避免迁移时因公共配置不同而需对算法进行修改。数据访问权限申请经审核通过,即可直连数据库或服务接口读取数据,直连数据库所需的表名和字段名可在天擎门户网站上查看。数据产品回存所需的存储目录或数据表,在产品注册完成后自动创建。监视信息须按照天镜的规范要求,通过接口发送至天镜。

4.2 算法镜像封装

算法在临时容器资源中完成开发与测试,在线

完成算法镜像一键封装,存入加工流水线的镜像仓库。算法镜像中包含算法实体、算法配置及算法依赖环境等,支撑算法的调度运行。

4.3 算法集成运行

算法集成运行包括算法注册、加载部署及任务定义。算法注册时需提交算法实体并填报算法元数据信息,经审核通过,加载部署实现算法与计算资源的关联。任务定义的核心操作是配置算法的调度策略,如针对定时调度,通过可视化 crontab 的方式设定算法的定时启动时间,针对数据源感知调度,需选定数据源(如中国地面自动站逐小时数据)并配置触发阈值(如每个时次到报 90%的站数)。完成任务定义,算法即可由流水线进行自动调度运行与生成产品。

5 业务应用效益

2021 年 12 月 15 日加工流水线随天擎开始全国业务运行,发挥业务效益:①实现算法统一管理,便于国省统一部署运维,同时基于加工流水线的数算一体、任务调度等技术,提升运行效率。②实现数据产品直存天擎,消除数据壁垒,有利于数据开放共享,且基于自然时序的数据源感知调度能够串联全业务链条,消除应用烟囱,提升各业务间的协同度。③建立“云+端”技术体制,初步实现“云+端”业务技术体制改革目标,构筑集约发展的基础。

截至 2023 年 11 月底,全国有 202 个应用系统融入天擎并基于加工流水线业务运行。国省融入气象算法 6739 个(图 8),气象算法库算法资源的规模效益初现,为国省算法共建共享提供基础。

国家气象中心的指挥决策分析系统(MESIS3.0)

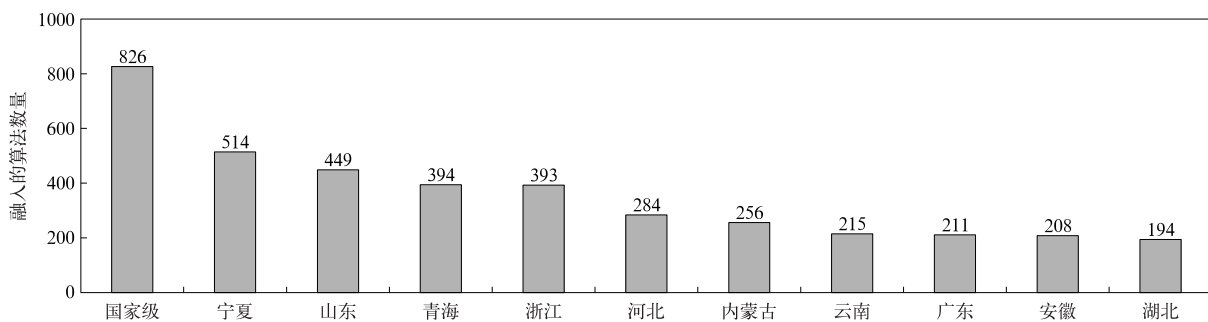


图 8 截至 2023 年 11 月底国省融入加工流水线的算法数量(省级取前 10 名)

Fig. 8 Number of meteorological algorithms integrated into data processing line from national-level and provincial-level by the end of Nov 2023(only the top 10 provinces are shown)

实现业务全流程深度融入天擎,40 余个算法基于加工流水线调度运行,支撑超 1000 个数据产品的高效生产,系统在多次天气过程中发挥重要作用。ME-SIS3.0 融入天擎后,集约化程度和系统性能大幅提升,其中图形产品加工效率由 15 s 提升至 5 s,极值统计效率提升近 10 倍。

国家气象信息中心基于位置的实况服务系统完全基于天擎原生设计开发^[41],为气象部门提供统一、准确、高时空分辨率的天气实况服务,2020 年 12 月业务运行,实时支撑气象实况微信小程序和中国气象 APP。融合站点分钟观测数据的实况格点产

品基于加工流水线实时生成,实现温度、降水、相对湿度、风等要素逐分钟和能见度、天气现象逐 5 min 快速更新。

中国气象局气象探测中心的天气雷达拼图系统 V3.0 于 2023 年 4 月开始基于天擎业务运行,实时生成单站和组网拼图产品 30 余种,提供国省用户应用。融入天擎后,应用加工流水线的数算一体和高效任务调度技术极大简化业务流程,减少数据搬运耗时,提升算法运行效率,雷达产品生成与服务时效由 6 min 提升至 1 min(图 9)。

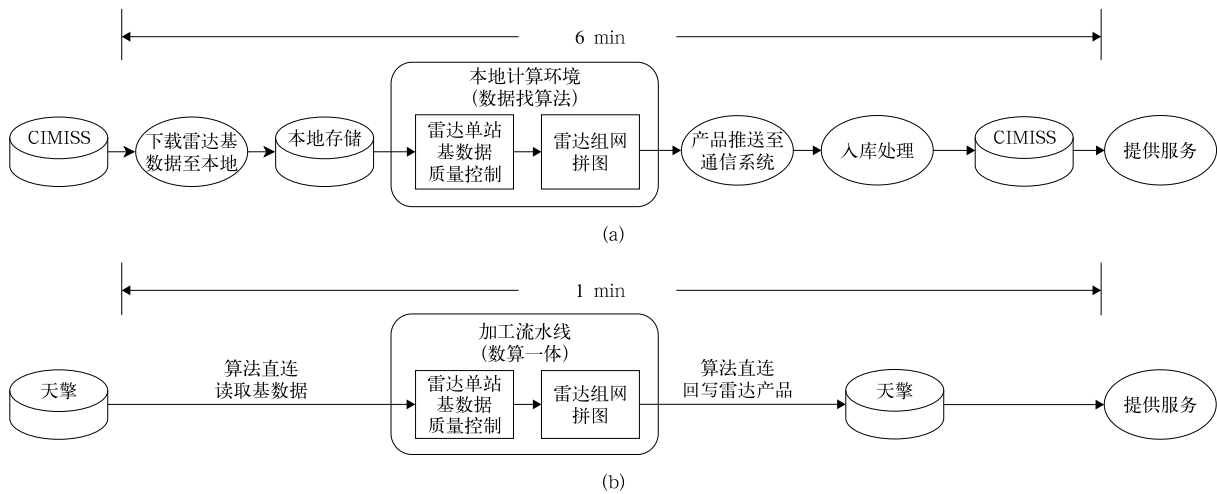


图 9 基于加工流水线的天气雷达拼图系统 V3.0 优化效果

(a)原流程,(b)现流程

Fig. 9 Optimization of weather radar mosaic system V3.0 based on data processing line

(a) the original, (b) the current

6 小 结

本文在分析业务需求的基础上,对加工流水线的总体架构、主要功能、关键技术等进行详细介绍,并总结系统的业务应用效益:

1) 通过建立算法库,实现国省气象算法统一管理,集成气象算法 6739 个,形成一定规模的气象算法资产,为气象算法的共建共享提供支撑。

2) 加工流水线建立数算一体、高效任务调度和加工流程可视化定义等技术,并充分应用容器技术,实现气象数据产品加工任务的高效、集约调度运行。

3) 国省两级业务应用证明加工流水线对提升气象业务系统的运行效率、消除数据壁垒和应用烟囱、建立“云+端”业务技术体制、推进气象业务集约

发展具有重要支撑作用。

随着大数据、云计算、人工智能等信息技术的不断发展,加工流水线需不断进行功能升级与性能提升:①应用 DevOps 思想实现流水线的敏捷开发与快速迭代升级,以快速响应多变的业务需求,减少系统升级对已运行业务的影响。②算法与资源调度需更加智能化,实现计算资源的动态扩缩容,提升资源利用效能。③建立完备的算法开发测试环境,支撑算法的高效开发、测试及向业务环境的一键迁移部署,提升算法业务上线的时效与规范性。

参 考 文 献

[1] 郑国光. 以信息化推进气象现代化. 浙江气象, 2015, 36(2): 1-4. Zheng G G. Promoting meteorological modernization by informatization. *J Zhejiang Meteor*, 2015, 36(2): 1-4.

- [2] 沈文海,赵芳,高华云,等. 国家级气象资料存储检索系统的建立. *应用气象学报*,2004,15(6):727-736.
Shen W H,Zhao F,Gao H Y, et al. The construction of national meteorological archival and retrieval system. *J Appl Meteor Sci*,2004,15(6):727-736.
- [3] 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. *应用气象学报*,2006,17(5):621-628.
Li J M, Shen W H, Wang G F. Meteorological information sharing service platform and its key technologies. *J Appl Meteor Sci*,2006,17(5):621-628.
- [4] 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. *应用气象学报*,2015,26(4):500-512.
Xiong A Y,Zhao F,Wang Y, et al. Design and implementation of China Integrated Meteorological Information Sharing System(CIMISS). *J Appl Meteor Sci*,2015,26(4):500-512.
- [5] 赵芳,熊安元,张小纓,等. 全国综合气象信息共享平台架构设计技术特征. *应用气象学报*,2017,28(6):750-758.
Zhao F,Xiong A Y,Zhang X Y, et al. Technical characteristics of the architecture design of China Integrated Meteorological Information Sharing System. *J Appl Meteor Sci*,2017,28(6):750-758.
- [6] 孙超,霍庆,任芝花,等. 地面气象资料统计处理系统设计与实现. *应用气象学报*,2018,29(5):630-640.
Sun C,Huo Q,Ren Z H, et al. Design and implementation of surface meteorological data statistical processing system. *J Appl Meteor Sci*,2018,29(5):630-640.
- [7] 吴焕萍,张永强,孙家民,等. 气候信息交互显示与分析平台(CIPAS)设计与实现. *应用气象学报*,2013,24(5):631-640.
Wu H P,Zhang Y Q,Sun J M, et al. Designing and implementation of Climate Interactive Plotting and Analysis System. *J Appl Meteor Sci*,2013,24(5):631-640.
- [8] 王若瞳,王建民,黄向东,等. MICAPS4 服务端系统架构设计. *应用气象学报*,2018,29(1):1-12.
Wang R T,Wang J M,Huang X D, et al. The architecture design of MICAPS4 server system. *J Appl Meteor Sci*,2018,29(1):1-12.
- [9] 韩丰,沃伟峰. SWAN2.0 系统的设计与实现. *应用气象学报*,2018,29(1):25-34.
Han F,Wo W F. Design and implementation of SWAN2.0 platform. *J Appl Meteor Sci*,2018,29(1):25-34.
- [10] 吴门新,庄立伟,侯英雨,等. 中国农业气象业务系统(CAgMSS)设计与实现. *应用气象学报*,2019,30(5):513-527.
Wu M X,Zhuang L W,Hou Y Y, et al. The design and implementation of China Agricultural Meteorological Service System(CAgMSS). *J Appl Meteor Sci*,2019,30(5):513-527.
- [11] 李德泉,李抗抗,李宏宇,等. 飞机作业监测移动应用系统的设计与实现. *应用气象学报*,2019,30(6):745-758.
Li D Q,Li K K,Li H Y, et al. Design and implementation of mobile application for real-time monitoring of weather-modification aircraft operations. *J Appl Meteor Sci*,2019,30(6):745-758.
- [12] 马强,颜京辉,魏敏,等. 北京气候中心 CMIP6 数据共享平台及应用. *应用气象学报*,2022,33(5):617-627.
Ma Q,Yan J H,Wei M, et al. Implementation and application of BCC CMIP6 Experimental Data Sharing Platform. *J Appl Meteor Sci*,2022,33(5):617-627.
- [13] 李莹,王国复. 气象灾害风险管理系统设计与应用. *应用气象学报*,2022,33(5):628-640.
Li Y,Wang G F. Design and implementation of Meteorological Disaster Risk Management System. *J Appl Meteor Sci*,2022,33(5):628-640.
- [14] 张进,孙健,沈学顺,等. CMA-GFS V4.0 模式关键技术研发和业务化. *应用气象学报*,2023,34(5):513-526.
Zhang J,Sun J,Shen X S, et al. Key model technologies of CMA-GFS V4.0 and application to operational forecast. *J Appl Meteor Sci*,2023,34(5):513-526.
- [15] 杨和平,张强,罗兵,等. 气象综合指挥平台建设与应用. *应用气象学报*,2023,34(1):117-128.
Yang H P,Zhang Q,Luo B, et al. Construction and application of Meteorological Integrated Command Platform. *J Appl Meteor Sci*,2023,34(1):117-128.
- [16] 梁丽,马舒庆,滕玉鹏,等. 天气雷达空中生态监测系统建设和应用. *应用气象学报*,2023,34(5):630-640.
Liang L, Ma S Q, Teng Y P, et al. Construction and application of Weather Radar Aerial Ecological Monitoring System. *J Appl Meteor Sci*,2023,34(5):630-640.
- [17] 沈文海. 气象信息化进程中云计算的意义. *中国信息化*,2015(3):80-88.
Shen W H. Significance of cloud computing in the process of meteorological informatization. *Chinese Informatization*,2015(3):80-88.
- [18] 沈文海. 气象业务信息系统未来基础架构探讨——“云计算”和“大数据”在气象信息化中的作用. *气象科技进展*,2015,5(3):64-66.
Shen W H. Discussion on the future infrastructure of meteorological business information system — The role of “cloud computing” and “big data” in meteorological informatization. *Adv Meteor Sci Tech*,2015,5(3):64-66.
- [19] 章国材. 气象云建设的研究与思考. *气象与环境科学*,2015,38(4):1-11.
Zhang G C. Study and thinking on the construction of meteorological cloud. *Meteor Environ Sci*,2015,38(4):1-11.
- [20] 聂峰英. 大数据资源技术服务协同研究——以气象数据为例. *信息化研究*,2016,42(1):6-11.
Nie F Y. The research of the big data resources technical services collaboration — A case study of meteorological data. *Informatization Res*,2016,42(1):6-11.
- [21] 沈文海. 对气象信息化的理解和再认识. *气象科技进展*,2013,3(5):56-62.
Shen W H. Understanding and re-understanding of meteorological informatization. *Adv Meteor Sci Tech*,2013,3(5):56-62.
- [22] 周勇,刘东君,马锋波. 气象信息化标准体系框架研究. *中国信息化*,2016(4):74-80.
Zhou Y,Liu D J, Ma F B. Research on the framework of meteorological informatization standard system. *Chinese Infor-*

- matization*, 2016(4):74-80.
- [23] 中国气象局. 气象大数据行动计划(2017—2020 年). 2017.
China Meteorological Administration. Plan of the Meteorological Big Data Action(2017—2020). 2017.
- [24] 中国气象局. 气象信息化发展规划(2018—2022 年). 2017.
China Meteorological Administration. Plan of the Meteorological Informatization Development(2018—2022). 2017.
- [25] 中国气象局. 中国气象局关于推进气象业务技术体制重点改革的意见. 2020.
China Meteorological Administration. Opinions of the China Meteorological Administration on Promoting the Key Reform of Meteorological Technology System. 2020.
- [26] 中国气象局. 新型气象业务技术体制改革方案(2022—2025 年). 2022.
China Meteorological Administration. Plan for the Reform of Meteorological Technology System(2022—2025). 2022.
- [27] 中国气象局. 气象大数据云平台设计方案. 2018.
China Meteorological Administration. Design of the CMA Big Data and Cloud Platform. 2018.
- [28] Papazoglou M P, van den Heuvel W J. Service oriented architectures: Approaches, technologies and research issues. *VLDB J*, 2007, 16(3):389-415.
- [29] Kivity A, Karrlay Y, Laor D, et al. KVM: The Linux Virtual Machine Monitor//Proceedings of Linux Symposium, 2007, 1: 225-230.
- [30] Kozhirbayev Z, Sinnott R O. A performance comparison of container-based technologies for the cloud. *Future Gener Comput Syst*, 2017, 68:175-182.
- [31] 杨鹏, 马志程, 彭博, 等. 集成 Docker 容器的 OpenStack 云平台性能研究. *计算机工程*, 2017, 43(8):26-31.
Yang P, Ma Z C, Peng B, et al. Performace research of Open-Stack cloud platform integrated with Docker container. *Comput Eng*, 2017, 43(8):26-31.
- [32] 伍阳. 基于 Docker 的虚拟化技术研究. *信息技术*, 2016, 40(1):121-123;128.
Wu Y. Research on virtualization technology based on Docker. *Inf Technol*, 2016, 40(1):121-123;128.
- [33] Salza P, Ferrucci F. Speed up genetic algorithms in the cloud using software containers. *Future Gener Comput Syst*, 2019, 92:276-289.
- [34] Wan X L, Guan X J, Wang T J, et al. Application deployment using Microservice and Docker containers: Framework and optimization. *J Netw Comput Appl*, 2018, 119:97-109.
- [35] De Benedictis M, Lioy A. Integrity verification of Docker containers for a lightweight cloud environment. *Future Gener Comput Syst*, 2019, 97:236-246.
- [36] Felter W, Ferreira A, Rajamony R, et al. An Updated Performance Comparison of Virtual Machines and Linux Containers//2015 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS). IEEE, 2015:171-172.
- [37] Joy A M. Performance Comparison between Linux Containers and Virtual Machines // Proceedings of 2015 International Conference on Advances in Computer Engineering and Applications. IEEE, 2015:342-346.
- [38] Merkel D. Docker: Lightweight Linux containers for consistent development and deployment. *Linux J*, 2014, 2014(239):No2.
- [39] Jaramillo D, Nguyen D V, Smart R. Leveraging Microservices Architecture by Using Docker Technology//Southeast Con. IEEE, 2016:1-5.
- [40] JGraph Ltd. User Manual(version 4. 2. 2-28). 2020.
- [41] 徐拥军, 倪学磊, 郑波, 等. 基于位置的天气实况数据服务接口设计与应用. *计算机系统应用*, 2023, 32(5):77-86.
Xu Y J, Ni X L, Zheng B, et al. Design and application of live weather data service interface based on user's location. *Comput Syst Appl*, 2023, 32(5):77-86.

Design and Application of Algorithm Intensive Environment for CMA Big Data and Cloud Platform

Huo Qing¹⁾ He Wenchun¹⁾ He Lin²⁾ Gao Feng¹⁾ Chen Shiwang¹⁾ Xu Yongjun¹⁾

¹⁾ (National Meteorological Information Center, Beijing 100081)

²⁾ (Shaanxi Meteorological Information Center, Xi'an 710014)

Abstract

With the advancement of meteorological services, various product processing systems and supporting data management systems have been developed for different business systems. However, this has led to the problem of system non-intensification. The lack of intensification in meteorological services can result in inconsistent data standards and make the operation and maintenance more challenging, which can lead to significant waste in investment due to duplicate data storage and inconsistent data caused by untimely synchronization. Moreover, the lack of information and technology hinders the integration of upstream and downstream businesses.

The intensive development of meteorological business systems and the reform of “cloud+end” technology system are important measures to achieve high-quality development of meteorological business. China Meteorological Administration proposed to build a “cloud+end” technology system in 2020. CMA Big Data and Cloud Platform serves as the cloud, while the meteorological business system is the end, clarifying the positioning of CMA Big Data and Cloud Platform as a key foundational technology platform. The data processing line (DPL) is an intensive environment for meteorological algorithms. It addresses business needs such as efficient and stable processing of data products, data sharing and collaboration among business systems, and efficient and intensive business applications. To achieve this goal, algorithm libraries and task control functions are established by utilizing technologies such as integrated digital and computing, efficient task scheduling, visual process arrangement, and containers. The algorithm library facilitates the standardization, unified management and sharing of algorithms. Task control supports multiple scheduling strategies with high reliability and fault tolerance, enabling efficient and stable scheduling operations for algorithms. All functions mentioned above are in the form of interfaces for the application frontend. At the same time, based on the meteorological business comprehensive monitoring system (referred to as Tianjing) enables automatic collection of algorithm operation status and detection of abnormal alarms. Since its operation in 2021, the processing assembly line has facilitated the real-time operation of 202 business systems nationwide, resulting in a performance improvement of 1–10 times and a significant increase in efficiency. It plays an important supporting role in improving the operational efficiency of business systems, enhancing their collaboration, accelerating the process of “cloud+end” business technology system reform, and promoting the intensive development of meteorological business.

Key words: data processing line; meteorology algorithm library; task scheduling; data and calculation integration; intensive development; “cloud+end” technical system reform