

基于 GIT 的气象数值模式代码管理平台 METCODE

赵春燕¹ 王彬^{1*} 孙婧¹ 常飏¹ 胡江凯² 周斌²

(1 国家气象信息中心, 北京 100081; 2 国家气象中心, 北京 100081)

摘要 随着气象数值模式的发展完善,已广泛应用于国家级和区域中心,研发力量地理上分散。支撑数值模式运行的高性能计算机系统呈现出异构众核的发展趋势,模式研发的学科交叉衍生性越发明显,分布式模式研发的高效协同共享成为一个亟待解决的问题。本文基于 Git 建立了一个代码管理协作共享平台 METCODE,实现了气象数值模式的代码版本管理、集成协作共享与过程管理,方便地支持分布式团队的合作研发。平台已在气象数值模式的研发中成功应用,应用效益良好。

关键词 METCODE;Git 版本控制系统;气象数值模式;代码管理

中图分类号: P409,TP319 **DOI:** 10.19517/j.1671-6345.20200442 **文献标识码:** A

引言

气象数值模式是气象科技创新的核心组成部分,已成为气象部门技术进步与业务发展的重要标志之一。气象数值模式不同于一般意义的软件产品,其研发是一个科学探索性较强的过程,同时需要规范严格的技术手段保障业务运行。由于兼具科研试验、业务管理以及科研—业务转化的多重需求,持续改进的过程管理和代码版本控制对气象数值模式的可持续发展至关重要。代码管理是软件配置管理中的基础管理过程^[1],是软件能力成熟度评估标准中的关键过程域。近几年,软件开发规模越来越大,开发团队也随之扩大,分布式开发成为常态,团队开发质量和协作效率问题也越来越突出,而优秀的代码管理平台能够支撑开发人员更高效地协同开发,同时保证软件资产的完整性和可跟踪性。

代码管理软件技术的发展经历了本地版本控制、集中式的版本控制和分布式版本控制 3 个阶段^[2]。在 20 世纪中叶,计算机得到初步应用,软件系统开发周期较长,出现了如 RCS 的基于计算机系统本地的代码管理工具^[3]。随着计算机的发展和软件规模的扩大,20 世纪 70、80 年代,出现了 C/S 架

构的集中式代码管理工具如 CVS、SVN、Perforce 等^[3],由服务端提供合并、版本记录集中式的管理。20 世纪 90 年代到 21 世纪初,随着互联网的蓬勃发展、软件规模日益扩大,分布式开发团队广泛出现,分布式版本控制系统应运而生,如 Git^[4]、Mercurial。随着互联网开放程度的提高与云计算技术的发展成熟,代码托管及共享服务平台、项目协作生态社区成为代码管理的主要形式,如全球最大的开源社交编程及代码托管网站的 GitHub^[5]、国内的码云 Gitee 和 CODING^[6],其分布式、高效协同共享的特点,使其成为现代代码管理的主要手段。国际上,欧洲中期天气预报中心、美国国家环境预报中心、美国国家大气研究中心均已采用 Git 分布式技术支撑模式协同研发,解决了国际合作研究的研发协同问题,提升了模式研发的协同效率。

气象数值模式^[7]是实现数值天气预报的核心技术手段,气象数值模式具有持续性、科学性、分布性、广泛交叉应用衍生等特性,对代码管理提出了新的挑战。以 GRAPES^[8](Global/Regional Assimilation PrEdiction System)模式为例,从 2001 年起至今,经历了近 20 年的持续改进升级,如何保持研发历史脉络的清晰和可持续性成为关键问题之一。

<http://www.qxkj.net.cn> 气象科技

“气候变化应对决策支撑系统”工程项目资助

作者简介:赵春燕,女,1984 年生,硕士,高级工程师,主要从事气象高性能计算应用支撑软件设计开发,Email:zhaocy@cma.gov.cn

收稿日期:2020 年 10 月 20 日;定稿日期:2021 年 2 月 26 日

* 通信作者,Email:wangbin@cma.gov.cn

GRAPES 模式包括资料同化^[9]、模式动力框架、物理过程等模块,基于 Fortran 采用模块化的设计^[10]组合调用,采用并行框架调试运行在气象高性能计算机系统^[11],融合了气象科学、计算科学,尤其在众核异构计算发展趋势下,支持众核加速计算架构下模式移植优化研发^[12]、跨学科跨部门的交叉协同合作成为迫切需求。GRAPES 模式已发展了 GRAPES_GFS^[13](全球中期预报模式)和 GRAPES_MESO(区域中尺度模式),各区域气象业务中心根据业务的特点发展了特色的应用模式,如集合预报^[14]、沙尘暴预报、台风预报^[15]、水文洪水预警预报等,拥有众多的分布式开发团队,研发力量地理上分散,涉及大量的分布式协同共享、集成合并管理需求。综上,需要建立支撑模式分布式协同众创研发的代码管理平台,支持大规模协作共享研发、科学规范代码集成管理和脉络清晰可追踪的研发成果管理。

本文围绕气象数值预报模式的研发特点和发展需求,通过技术选型分析,设计建立了一个基于 Git 的分布式代码协作共享管理平台 METCODE,保存模式研发历史,支撑气象数值模式的持续发展和国省合作研发应用的协作共享,建立了模式代码管理标准,规范化研发、集成合作和业务化各个环节。

1 气象数值模式研发管理现状

2010 年,中国气象局引进了商业版本管理软件应用于 GRAPES 研发^[16],建立了气象数值模式代码管理系统,首次实现了 GRAPES 模式研发的代码管理,运行至今建立了 956 个用户终端,更新次数 21260 次,系统集成 2059 次,支撑发布版本 21 套。随着模式研发历史的累积,系统访问性能开始出现下降。同时,由于服务端软件版本没有升级,一定程度上限制了客户端软件版本的兼容性,在新的高性能计算平台上应用存在困难。

发展我国自主可控的气象数值模式,是一个持续发展的科学工程。气象科学发展迈入了地球系统时代,天气和气候模式呈现一体化发展融合的趋势。天气、环境、水文等学科呈现更广泛交叉应用衍生的趋势,模式研发的生态逐渐建立,模式研究更加开放共享,分布式协作的需求更加突出。因此,亟待建立一套安全、分布、支持协作共享的代码管理系统。同时,随着地球系统模式发展、数值模式预报的精度与

预报时效的不断提高,科学计算规模呈指数级增加,异构众核、低耗高效超算架构体系^[17]成为高性能计算的发展趋势,多态、异构^[12]的研发计算平台需要代码管理工具具备较高的兼容性,广泛的支持混合计算架构^[13]下的研发交互以及结合代码编辑工具的集成使用。

2 METCODE 平台设计

气象数值模式代码管理平台 METCODE 在 Git、Docker、Spring、React.js、MySQL、Redis 等开源软件基础上,进行了自主定制开发形成了国产代码管理和共享协作平台,部署于中国气象局私有云^[18-19],是以 B/S 为主、C/S 为辅的整套应用级软件,支持 DevOps^[20]研发运维一体化的过程管理。

2.1 平台功能

METCODE 平台由云端代码管理共享协作平台和云端集成开发工具组成(图 1)。云端代码管理共享协作平台以 Web 方式,围绕研发流程提供项目管理、分布式代码管理、文件管理、知识管理及 DevOps 等功能,以项目为组织单元建立代码库,实现对项目的代码管理和协作共享,支持对项目人员及权限管理、集成共享管理、项目文件、项目任务、版本发布等进行管理。云端集成开发环境支撑协同编辑的 Web 在线开发终端环境,定制了气象主要编程语言 Fortran 语法提示功能。

METCODE 代码管理共享协作平台支持实现项目生命周期的 DevOps 一体化全过程管理,可用于气象各类项目建设的过程管理和成果管理,包含项目管理、任务管理、代码管理、持续集成、测试管理、部署管理和构建管理等功能,实现了结合云端开发环境提供代码的在线开发、编译、部署。

2.2 气象数值模式研发应用设计

METCODE 基于中国气象局私有云部署,根据用户分布式的特点,设计支持气象部门内部用户、区域中心的用户和外部用户的合作研发管理。设立了备份系统,对关键数据进行定时备份以保证应用数据的安全完整。METCODE 提供统一域名访问 Web 页面,支持项目管理和代码管理、集成交互和共享,发挥代码托管和共享的主要功能(B/S 方式),研发人员可在个人电脑、笔记本电脑上通过浏览器访问 Web 网页界面或通过可视化客户端来访问 METCODE 代码库。气象数值模式一般基于高性能

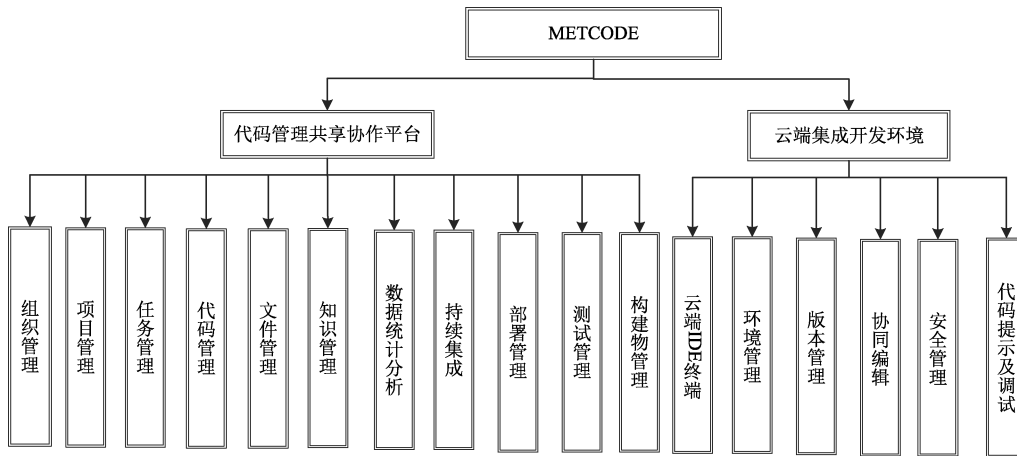


图 1 METCODE 功能结构

计算机系统研发和业务运行,在中国气象局“派-曙光”高性能计算机上部署 Git 客户端,通过命令行操作可与 METCODE 代码库交互,开展模式研发、模式异构众核移植优化的代码管理(C/S 方式)。用户

也可通过 Git 可视化客户端、集成 IDE 结合 METCODE 域名,实现代码管理。为了支持内外部用户合作研发成果的集成共享,平台实现了代码库分支级别的数据同步功能(图 2)。

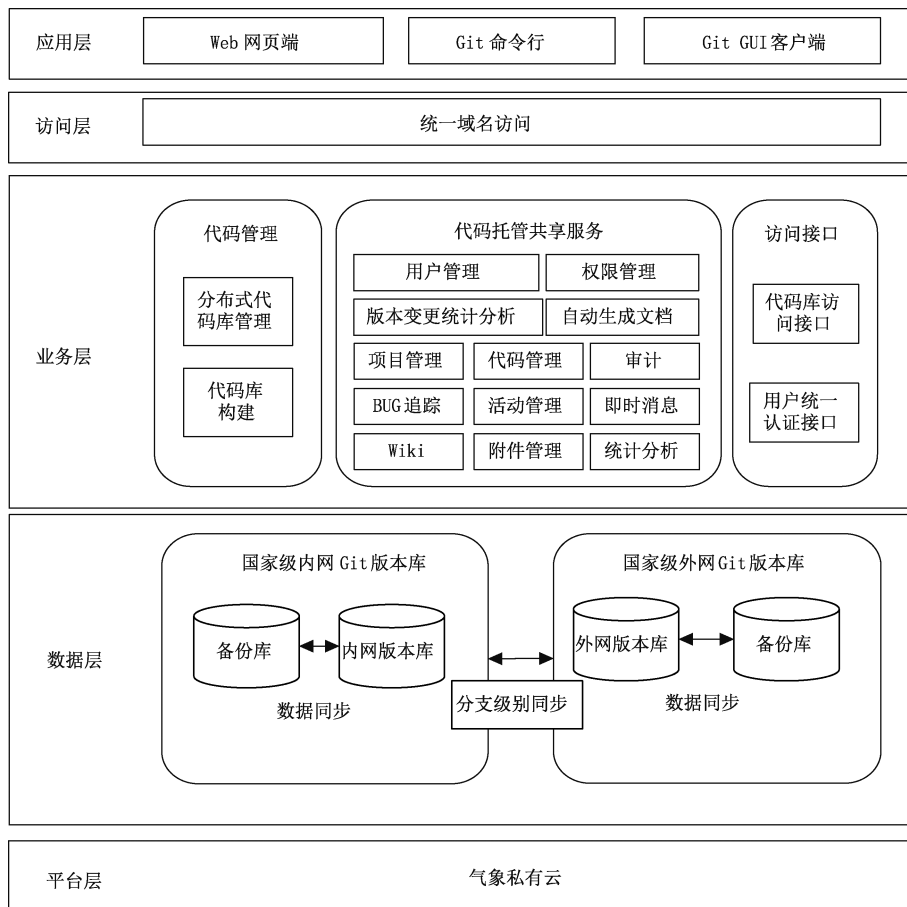


图 2 气象数值模式研发应用示意

模式研发用户在“派-曙光”高性能计算机上通过命令行访问 METCODE 代码仓库的工作流程如图 3 所示,首先从远程代码库克隆到高性能计算机本地工作目录,或者从远程代码库拉取代码到本地代码库,并与本地代码库进行合并。用户在本地用户工作目录下进行研发(增、删、改),将修改的版本提交到本地代码库,文件先被添加到缓存区,再被提交到本地代码库,通过推送到 METCODE 远程代码库统一管理。

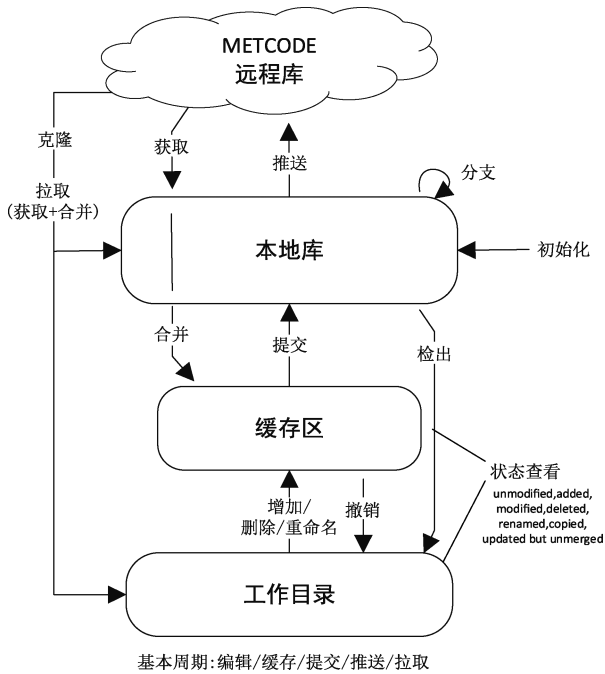


图 3 “派-曙光”访问流程

3 关键技术及实现

3.1 平台技术架构

METCODE 是 B/S 架构的系统,基于国产 CODING 软件系统定制开发,采用 Linux、容器技术,基于微服务^[21]、服务发现方式进行架构,具有去中心管理、粒度小、耦合度低、易扩展、易维护等优点。系统从上至下分为应用层、微服务层、数据库层和基础架构层(图 4)。基础架构层采用容器技术编排维护系统,数据库层采用 MySQL 关系型数据库存储项目等通用信息,采用 Redis 缓存技术实现数据的高效访问。微服务层采用了 Git-Server 和 Repo Manager 实现代码仓库的传输服务和管理,将代码库的管理、维护、备份等功能通过微服务组件管理,以精细粒度管理应用,耦合了 Git 本身可独立应用的特点。基于 SSO 技术实现了用户统一认证和单点登录集成,应用层采用 Nginx 实现服务请求管理,并实现了分布式架构的负载均衡,基于云架构可有效支持高并发访问,集成了集成开发环境插件及第三方的 Jenkins、Sonar 等插件。

3.2 用户统一认证及单点登录

“派-曙光”高性能计算机系统设计有一套规范的高性能计算机用户管理机制,并建有北京高性能计算机应用中心门户网站,实现了用户的统一管理和单点登录。为了便捷支撑研发人员在“派-曙光”上基于 METCODE 开展模式研发协同,实现了 METCODE 与北京高性能计算机应用中心的用户

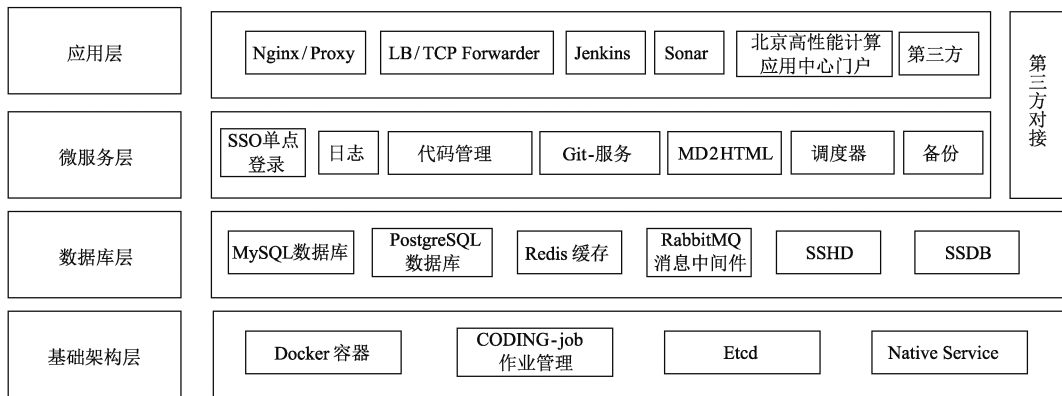


图 4 METCODE 技术架构

统一认证管理和单点登录^[22],采用 SSO 认证机制,与北京高性能计算机应用中心 SSO 认证中心的接口通信以交换令牌、校验令牌及发起登录、注销请

求。认证流程如图 5 所示,用户通过向门户网站发送登录请求,门户网站将该请求转发到统一认证中心进行认证请求,认证中心认证通过后,返回认证唯

一令牌到门户网站,门户网站携带认证中心提供的令牌信息,向被保护资源发送访问请求响应。

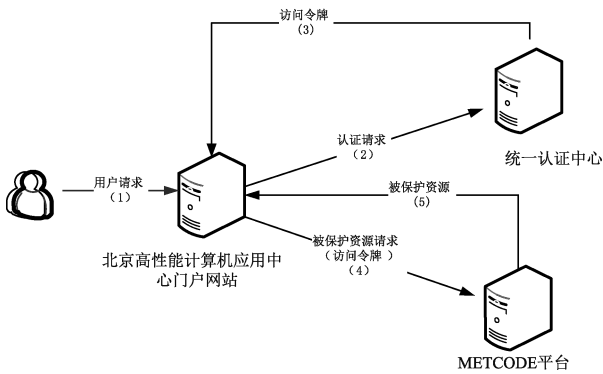


图 5 METCODE 用户统一认证流程

3.3 分支同步技术

国家级气象数值模式研发主要基于中国气象局骨干网进行,解决了国省合作、外部单位的合作研发协同,保障模式代码库的安全性,实现了骨干网代码库和互联网 DMZ 区代码库分支级别的同步功能,支持内外网研发成果交互共享。分支同步由内网项目管理员发起,支持双向分支同步,基于 SSH 协议和接口技术实现。通过 Git 获取访问分支的公钥,读取分支内容,基于 SSH 协议传输至目标系统,再通过 Git 操作提交到目标项目中,实现分支信息的同步(图 6)。

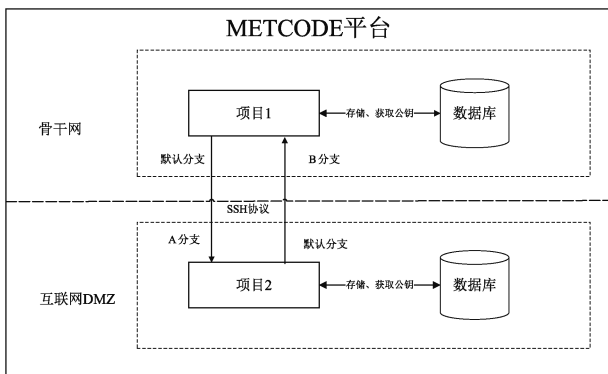


图 6 METCODE 分支同步示意

4 业务应用

METCODE 在气象部门内得到成功应用,支撑了 GRAPES 全球/区域模式、GRAPES 大气化学耦合模式(GRAPES_CUACE)联合研发,为气候海洋模式(MOM5)、模式后处理^[23]、模式 GPU 移植开

发、模式支撑软件、气象信息软件等项目建立了协同开发和代码管理功能。

4.1 代码库迁移

为了保障模式研发历史的延续性,实现新旧系统的无缝过渡迁移,在 METCODE 建立了 GRAPES 模式的历史代码库,用于迁移在原有代码库中的模式研发历史。在不同技术架构的代码库中,完整的迁移研发历史信息,需从原系统迁出历史记录,导入到新的系统保存。基于 Git-p4 迁移工具进行了二次开发,调用原有代码库命令行工具读取原代码库的内容,包括用户信息、路径信息、文件类型信息、变更信息等,通过 METCODE 的命令行工具将读取出的内容推送到 METCODE 的代码库中。对原有代码库中的路径信息、文件名、作者信息、代码正文中的中文进行过滤(路径、文件名、作者信息、文件内容),对 .docx, .doc, .pdf 等二进制文件进行过滤、正文非转码处理,实现了中文 UTF8 转码,解决了 GRAPES 模式研发历史迁移中文乱码的关键问题,完成了 GRAPES 历史库近 10 年超过 10 GB 研发历史数据的迁移(图 7)。

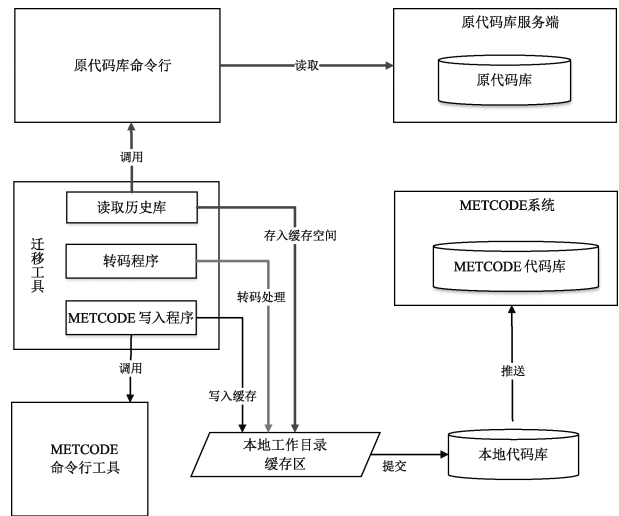


图 7 模式研发历史迁移原理

原有模式代码管理系统为 C/S 架构,模式的研发历史均存储于服务端元数据库,而基于需求分析中的统计数字,集成更新次数只占总更新次数的 10%左右,而约 90%的研发历史记录是团队内部及个人的研发记录,大量的元数据必然导致系统性能降低,同时也不利于保持模式研发历史脉络的清晰。基于 METCODE 代码管理平台,利用 Git 的分布式

概念,将个人研发记录保存在本地库, METCODE 远程库仅保存团队以上的研发成果和历史,能够减少大量个人研发记录,优化了代码库的管理,更好地支撑模式持续发展。

4.2 系统性能

基于生产环境采用压力测试工具模拟 200 并发

数对 METCODE 的任务列表访问、任务创建、代码库访问、代码目录浏览、代码源文件获取等关键应用场景的指令执行、响应时间、网络吞吐和网络收发速度进行了压力测试。表 1 测试结果显示各类操作响应时间平均小于 5 s,性能满足了模式分布式研发的需求。

表 1 METCODE 性能测试

项目请求	执行		响应时间/ms						吞吐量	网络/(KB·s ⁻¹)	
	线程数量	失败率/%	平均耗时	最小耗时	最大耗时	90%线程耗时	95%线程耗时	99%线程耗时	每秒钟交易量	接收	发送
数据库访问	200	0.00	4236.47	1227	12273	6453.00	7416.40	10156.21	12.14	26.14	2.71
任务创建	200	0.00	5996.82	1512	13692	11164.10	12566.25	13657.15	10.04	26.30	3.56
源文件获取	200	0.00	4303.80	897	7519	6275.10	6844.70	7270.22	13.04	12.19	2.62
任务列表访问	200	0.00	2981.40	499	9295	4753.40	5734.85	8433.81	9.30	205.32	1.92
代码目录浏览	200	0.00	3613.12	1233	6206	5229.60	5505.95	6016.50	19.35	85.94	4.03
代码库访问	200	0.00	3818.76	1086	7916	5489.90	5718.65	6817.70	24.32	141.73	5.15
总计	1200	0.00	4158.40	499	13692	6461.30	7591.65	12273.00	33.07	209.75	7.75

5 结论

在气象数值模式发展的需求驱动下,设计建立了基于 Git 的国有自主知识产权 METCODE 代码管理协作共享平台。METCODE 采用高可靠的架构设计、分布式的部署,能够安全、灵活地支撑内部、外部及合作研发的需求。METCODE 现已支持了 100 多个项目的管理,应用单位覆盖国家气象中心、国家气候中心、中国气象科学研究院、国家气象信息中心等业务科研单位,以及部分合作单位。

目前系统在企业级管理方面还需加强,如对多层次单位组织架构的支持和分级管理,将能更好地支撑中国气象局各类大型、中型和小型等各类项目的灵活管理。METCODE 随着 Git 的深入发展具有更加广泛的应用前景,未来将进一步完善,支撑全部气象数值模式的研发管理,并推广应用到气象软件系统项目的代码与过程管理。

参考文献

[1] 刘燕秋, 勉玉静, 赵文耘. 软件配置管理中版本管理技术研究[J]. 计算机工程与应用, 2003, 39(21): 68-71.
 [2] 赵春燕, 孙婧, 魏敏. 云及高性能计算集群环境中配置管理系统设计[J]. 计算技术与自动化, 2016, 35(1): 111-116.
 [3] 刘冉, 肖然, 覃宇. 代码管理核心技术及实践[M]. 北京: 电子工业出版社, 2018.

[4] Git 介绍 [EB/OL]. [2021-01-25]. <https://git-scm.com/>.
 [5] Github 介绍 [EB/OL]. [2021-01-25]. <https://github.com/>.
 [6] CODING 介绍 [EB/OL]. [2021-01-25]. <https://coding.net/>.
 [7] Lorenz E N. Reflections on the conception, birth, and childhood of numerical weather prediction [J]. Annual Review of Earth & Planetary Sciences, 2006, 34(1): 37-45.
 [8] 陈德辉, 薛纪善, 沈学顺, 等. 我国自主研发的全球/区域一体化数值天气预报系统 GRAPES 的应用与展望[J]. 中国工程科学, 2012, 14(9): 46-54.
 [9] 薛纪善, 庄世宇, 朱国富, 等. GRAPES 新一代全球/区域变分同化系统研究[J]. 科学通报, 2008, 53(20): 2408-2417.
 [10] 薛纪善, 陈德辉. 数值预报系统在 GRAPES 的科学设计与应用 [M]. 北京: 科学出版社, 2008.
 [11] 王彬. 高性能计算技术在气象部门的应用[J]. 计算机工程与设计, 2014, 35(4): 1476-1479.
 [12] 肖洒, 魏敏. BCC_AGCM 大气环流模式异构众核加速技术 [J]. 气象科技, 2018, 46(2): 511-517.
 [13] 沈学顺, 苏勇, 胡江林, 等. GRAPES_GFS 全球中期预报系统的研发和业务化[J]. 应用气象学报, 2017, 28(1): 1-10.
 [14] 霍振华, 陈静, 李晓莉, 等. GRAPES 业务全球集合控制预报初值的动力升尺度方法[J]. 气象科技, 2018, 46(4): 707-717.
 [15] 吴俞, 麻素红, 李勋, 等. GRAPES_TYM 模式对 TC 路径及环境引导气流预报检验分析[J]. 气象科技, 2016, 44(6): 937-948.
 [16] 赵春燕. GRAPES 版本管理系统备份恢复机制研究与实现 [C]//2011 年中国气象学会气象通信与信息技术委员会暨国家气象信息中心科技年会, 2011: 132-137.
 [17] 王彬, 孙婧. 气象高性能计算系统的业务发展概述[J]. 气象科

- 技进展,2018,8(1):287-289.
- [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[J]. 气象科技进展,2015,5(3):64-66.
- [19] 王彬,韩同欣,李楠. 气象私有云环境下存储架构设计与性能分析[J]. 计算机技术与发展,2017,27(5):20-24+29.
- [20] 耿泉峰,李曦,葛维,等. 基于 DevOps 的软件开发管理模式[J]. 软件,2019,40(1):93-96.
- [21] 郝江波,唐卫,王慕华,等. 基于微服务的气象信息决策支撑系统重构与实践[J]. 气象科技,2020,48(6):829-835.
- [22] 王彬,孙婧,常飏,等. 面向气象业务整合协同的信息共享门户系统[J]. 计算机应用研究,2013,25(增刊):2546-2549.
- [23] 佟华,胡江林,张玉涛. GRAPES 模式后处理技术改进应用研究[J]. 气象科技,2020,48(4):511-517.

METCODE: A Git-Based Code Management Platform for Meteorological Numerical Model

ZHAO Chunyan¹ WANG Bin¹ SUN Jing¹ CHANG Biao¹ HU Jiangkai² ZHOU Bin²

(1 National Meteorological Information Center, Beijing 100081; 2 National Meteorological Center, Beijing 100081)

Abstract: Meteorological numerical models are the core component of meteorological science and technology innovation, and have become one of the important symbols of technical development progress of meteorological departments. Unlike software products in general, the development of meteorological numerical models is a scientific exploration process and requires strict technical means to ensure operational running. Due to multiple requirements of scientific research experiment, operation management and research-operation transformation, the continuous improvement process management and code version control are crucially important to the sustainable development of meteorological numerical models. With the development and improvement of meteorological numerical models, it has been widely used in national and regional centers by the geographically dispersed research and development personnel. The development trend of heterogeneous multicore can be seen in the field of high-performance computing, and the interdisciplinary derivation of model research and development becomes more obvious, so the efficient collaborative sharing of distributed model research and development efforts becomes a key problem. In this paper, the METCODE platform is established based on the Git version control system which provides code management and collaboration services. The platform realizes code version management, integrated collaborative sharing and process management of meteorological numerical models, and supports the cooperative research and development of distributed teams conveniently. The platform has been successfully applied in the research and development of the GRAPES model with good results.

Keywords: METCODE; Git version control system; meteorological numerical models; code management