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.
Huang Zhi,Huang Heng,Liang Weiliang,et al. Design and application of business integration based on CMADaaS DPL[J]. Journal of Meteorological Research and Application,2022,43(1):73-77.

基于“天擎”DPL 的业务融入设计与应用初探

黄 志[1], 黄 珩[2*], 梁维亮[3], 梁苑苑[1]

(1.广西壮族自治区气象信息中心, 南宁 530022; 2.河池市气象局, 广西 河池 547000;
3.广西壮族自治区气象台, 南宁 530022)

摘要: 依据“天擎”应用融入标准和业务规范,以气象短时临近强降水预警服务处理流程为例,给出了基于“天擎”DPL 加工流水线的业务迁移融入设计流程,对应用融入账户注册、公共元数据注册、数据存储设计、DPL 处理算法设计与部署,以及通过“天镜”呈现 DPL 算法运行监控等关键环节进行了描述,为气象业务系统向“天擎”迁移和开展“云+端”融入改造提供技术思路。

关键词: 天擎;云+端;DPL;业务融入

中图分类号: P49 **文献标识码:** A **doi:** 10.19849/j.cnki.CN45-1356/P.2022.1.13

引言

随着云计算、大数据等新型技术的发展与应用,气象部门顺应技术发展趋势, 运用云存储与云计算平台以及大数据技术,实时收集、处理、存储和分析海量气象大数据,以分布式节点存储、多节点并行计算等方式提供智能、可线性扩容的气象大数据处理能力,建立统一的气象大数据环境,构建开放、智能的众创业务平台和监控平台,有效提升服务支撑能力,让数据和计算能力成为普惠社会和经济的基础[1],气象大数据云平台“天擎”(CMADaaS)作为全国气象部门信息化的核心平台应运而生。

“天擎” 作为全国气象部门信息化的核心平台,上线后将替代 CIMISS 作为国省新一代统一数据环境,助力气象业务高质量、协同创新发展[2-3]。中国气象局正在积极推进气象业务技术体制改革, 并明确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”的气象技术体制,实现数据管理、加工处理、应用服务的高度集约;推进气象业务系统与气象大数据云平台的深度融合, 实现各项业务和系统之间有效衔接和有机互动。

“天擎”作为全国气象部门信息化“数算一体”的核心平台,是促进形成“云+端”业务模式新格局的核心软件技术平台, 是消除 “数据孤岛”、“应用烟囱”,推进资源整合、流程再造,发挥数据聚集规模效应, 实现业务集约高效的关键。2020 年 12 月,“天擎”1.0 完成全国部署并进入业务试运行阶段, 形成了“1+31”的业务布局;山东省率先开展了对“天擎”平台部署和业务试运行[4],青海省就“天擎”的数据接口与 CIMISS 系统的 MUSIC 接口进行了比对分析,论述了“天擎”不仅解决 CIMISS 服务接口现状问题, 而且降低存储技术开发难度, 解耦应用与数据,直接支撑业务系统,能与现有业务系统实现平稳衔接[5]。广西气象部门相关业务单位如广西壮族自治区气象信息中心、广西壮族自治区气象灾害防御技术中心也率先对其所辖重要业务系统开展基于“天擎”的融入再造工作,并取得了初步的成效。本文以气象短时临近“三小时”强降水精细化气象预警服务处理流程为例(以下简称“三小时”预警服务),针对其业务处理流程从数据注册、数据存储、处理流

收稿日期: 2021-12-07
基金项目: 2021 年广西气象科研计划指令性项目(2021ZL02)
作者简介: 黄志(1981—),男,硕士,高级工程师,主要研究方向:气象信息化、大数据、云计算等。E-mail:616646373@qq.com
通讯作者: 黄珩(1984—),男,工程师,主要从事气象信息化、大数据应用研究。E-mail:361121687@qq.com

程再造、算法融入“天擎”加工流水线以及“天境”运行监控全过程设计，探讨各相关关键环节的设计和实现方法，为今后业务系统基于“天擎”的“云+端”服务模式改造融入提供技术支撑和解决思路。

1 “天擎”DPL 应用流程概述

“天擎”继承了 CIMISS 的全部标准规范，优化数据汇聚、加工、服务和应用支撑全流程，全面提升数据交换和存储服务时效。它的一个显著的功能就是提供“数算一体”的平台化服务，具备了全面支撑“云+端”的气象业务的能力。其中“天擎”的 DPL (Data Processing Line)加工流水线子系统，是“天擎”的一个特色模块，它以提升气象数据的加工处理能力、优化数据加工处理流程为主要目标，是进行“天擎”业务应用系统剥离算法、算法上云，产品直接云上存储，进行云化改造的核心功能模块。“天擎”DPL 中的“算法”是指为解决一个特定问题、生成一类数据产品或实现一个业务逻辑所定义的有序指令集。包括源代码和可执行文件两种形态，其中源代码供接口开放平台和众创支撑平台封装使用，可执行文件供加工流水线加载使用；DPL 的“加工处理任务”是针对相对静态的算法添加了一定的策略并在加工流水线上动态执行，称为发布一个加工处理任务。所以“天擎”的 DPL 能够实现某个加工处理任务在流水线上运行前进行算法部署、资源分配、执行策略配置等相关功能，以及解决所有任务运行过程中的调度问题，并具备对整个加工流水线的任务进行综合监视和控制的能力。今后在业务系统融入“天擎”的过程中，对业务系统的各个业务处理流程进行分解，并改造成适合“天擎”DPL 运行的加工流水线，是业务系统进行“天擎”融入改造的关键环节，并助力“天擎”最终实现数据管理、加工处理、应用服务的高度集约。个人或业务系统模块算法的“天擎”DPL 设计与融入流程如表 1 所示。

表 1 “天擎”DPL 设计与融入流程

步骤	步骤名称	步骤简单说明
步骤 1	“天擎”应用融入账户注册	个人通过“气政通”实名认证注册“天擎”应用融入账户，获得数据存储、接口访问、公共元数据等资源和权限
步骤 2	数据注册、存储设计、接口发布	数据产品四级编码注册、遵循对应的存储标准、库表设计和匹配 MUSIC 接口
步骤 3	DPL 加工流水线设计与部署	基于“天擎”MUSIC 接口或数据库直连方式，采用 Java、Python 等开发语言设计数据处理流程，封装后部署于“天擎”DPL 加工流水线
步骤 4	DPL 运行监控接入“天镜”	融入“天擎”的 DPL 加工流水线需要按照“天镜”的标准向其发送 DI/EI 信息，并通过“天镜”实现 DPL 运行监视

2 “三小时”预警服务业务需求与现状

“三小时”预警服务是气象短时临近预报的重要业务，该服务根据“三小时”强降水精细化气象预警工作流程和服务标准，视降雨发生发展情况逐小时或逐 3h 开展一次精细化服务或暴雨预警服务。该服务主要启动标准是通过实时统计全区自动站的 1h、3h、6h、24h 累计降水量，如超过规定的阈值和该站点的历史极值，则发布的预警信息。该服务(算法)的区级降水阈值判断标准如下：

(1)全区范围内有 1 个自动气象站前 1h 雨量≥80mm，如果有站点 1h 雨量实况突破当地历史极值的。

(2)全区范围内有 1 个自动气象站前 3h 雨量≥100mm，如果有站点 3h 雨量实况≥200mm(沿海站点≥300mm)。

(3)全区范围内有 1 个自动气象站前 6h 雨量≥150mm，如果有站点 6h 雨量实况≥250mm(沿海站点≥350mm)。

(4)全区范围内有 1 个自动气象站前 24h 雨量≥250mm，如果有站点 24h 雨量实况≥300mm(沿海站点≥400mm)。

该服务原处理流程是外挂单机版的应用程序，通过读取 CIMISS 的地面小时数据进行累加和阈值判断，结果回写外置数据库。本服务原处理流程如图 1 所示。

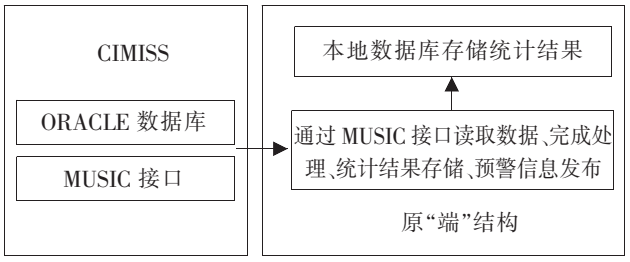

图 1　三小时预警服务原处理流程图

3 “三小时”预警服务业 DPL 融入设计与实现

3.1 “天擎”应用融入账户注册

“天擎”实行“气政通”个人实名认证登录，个人用户可以通过 OA 账户自动登录 “天擎”PORTAL。“天擎”的账户按照功能分为两类：MUSIC 接口账户和应用融入账户。MUSIC 接口账户用于通过接口调取气象数据，应用融入账户用于进行系统融入改造。按照气象大数据云平台业务管理规定，账户使用分两种情况：业务系统使用、个人科研使用，分别有不同的管理要求。

“天擎”开放的资源有数据（数据存储、数据访问、数据管理、数据归档、公共元数据等）、算法（算法计算资源，算法库管理，产品加工调度等）、接口（数据服务接口、接口众创服务、接口管理与发布等）。本服务需要进行 DPL 业务流程再造，在此过程中会涉及到产品数据四级编码注册、数据存储、MUSIC 接口的应用以及天镜的运行监控。所以需要注册“天擎”应用融入账户，才能拥有在“天擎”规范下进行数据公共元注册（MMD）、数据存储设计（SOD）、MUSIC 数据接口发布、匹配的算力资源等功能权限，实现对数据产品在“天擎”环境下的数据融入，并进行后续的 DPL 加工流水线的设计和部署运行。

3.2 “天擎”业务数据注册

“天擎”遵循和继承了 CIMISS 资料的四级编号规范，气象资料四级编码规则为X.YYYY.NNNN.MZZZ，第 1 位“X”大写字母标识 14 大类资料的简码；第 2 级“YYYY”资料编码 4 位数字标识国家级和省级编码分段（广西为“6800-6999”），目前已初步固化分配至各单位。第 3 级”NNNN” 资料编码 4 位数为 2 级资料下面的小类，作为各个单位自身各业务系统的分类，由广西壮族自治区气象局各直属单位和市气象局为自己的业务系统自定义编码段。

确定了四级编码之后，应用融入账户需要完成该数据产品四级编码的公共元数据（MMD）注册，在注册的过程中完成拟发布的数据产品分类、数据产品表结构设计方案的提交[6-10]和所需的 MUSIC 接口；在 MMD 审核通过之后，存储管理模块（SOD）会将提交的表结构的数据库字段与 MUSIC 接口字段进行映射绑定，并自动完成物理表的创建和对应的 MUSIC 接口发布。图 2 为“天擎”环境下数据注册发布流程图。

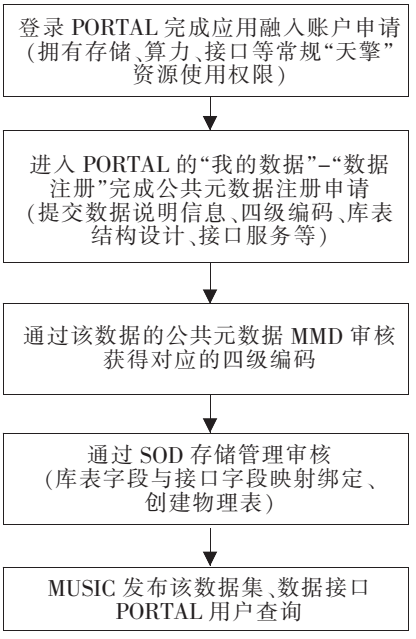

图 2　“天擎”业务数据注册流程

3.3 DPL 处理流程设计与部署

MUSIC 是用户从“天擎”数据环境中获取数据或产品的应用编程接口，采用瘦客户端模式。MUSIC API 支持 windows、Linux 以及 AIX 操作系统，包含常用的开发语言工具包主要有.NET、Java、PHP、C/C++、Python 等，其他语言开发包将根据用户需求给予支持。本算法基于“天擎”提供的 Java 客户端开发包，客户端开发包由一组库文件（或 jar 包）加配置信息（Client.config）组成，通过加载客户端库文件即可调用接口函数，实现从“天擎”数据环境中获取数据或产品功能，用户在进行应用程序开发时只需要完成算法逻辑流程的程序编写以及最终的程序封装。

本服务数据处理、数据存储以后最终预警产品获取可按“天擎”的“云+端”融入改造要求进行“云”端业务融入改造，而预警信息发布流程“端”则置于“天擎”环境外运行。“天擎”环境下的 DPL 改造融入

流程如下图 3 所示。

目前 DPL 只支持可执行的运行程序,所以用户提交的算法必须以可执行的形态运行[1]。落实到具体操作部署,即个人业务账户在“天擎”PORTAL“我的算法” 模块下对封装好的 jar 包在线进行 DPL 算法注册,并提交至 DPL 管理后台经管理员对该算法进行审核生效后,个人用户加载该算法至 DPL 集群相关运算节点对应的权限运行目录, 并通过设置定时触发的形式来启动该算法加工流水线的运行。

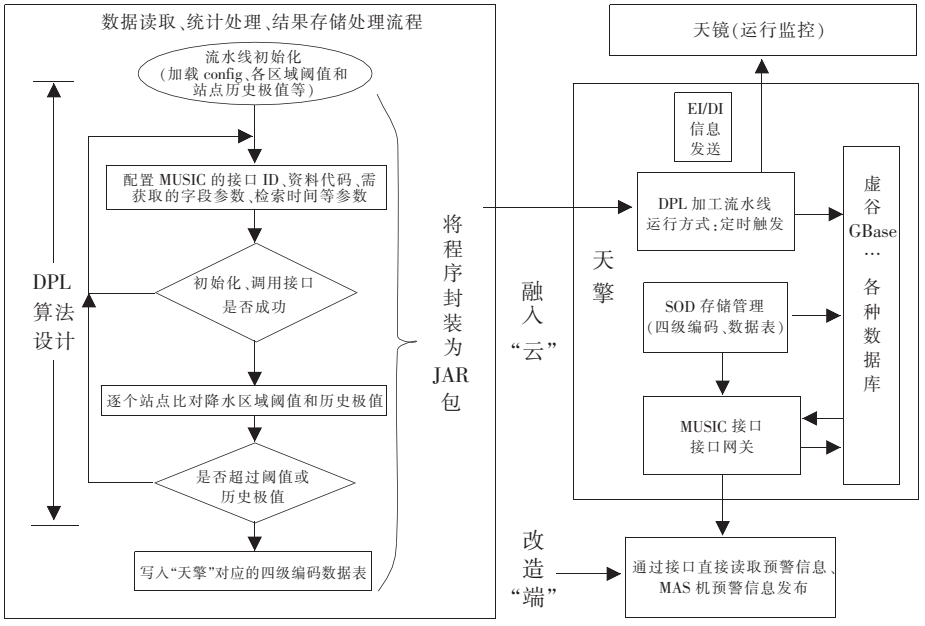

图 3 “天擎”环境下的 DPL 改造融入流程

个人应用融入账户基于 DPL 模块可以实现个人算法的有效管理(算法分类存储、版本管理、检索下载、授权使用、运行管理等),同时可申请他人算法的使用权限;用户在具备算法使用权限(本人算法或已被他人授权的算法)的基础上,使用任务定义功能可以创建自己的加工处理任务, 并可以配置任务的调度策略,进一步通过任务管理功能,可以对任务进行启停操作,生成自己的产品[1]。

3.4 基于“天镜”的 DPL 运行监控

目前“天擎”DPL 的业务运行监控环节,依据天镜标准接口改造升级所有加工处理流程, 增加发送 DI 和 EI 信息的业务功能, 并通过接口的形式对接天镜系统, 并在天镜显示其运行状况。DPL 的算法管理 Web 前端对接“天擎”个人“我的算法”管理接口,所以个人账户在“天擎”PORTAL 通过“我的算法”管理页面,可直观的看到 DPL 算法任务各个环节的调度计划以及下次执行时间,同时通过“日志查看”可以看到各环节执行时的日志输出。

经本算法统计处理的结果通过 MUSIC 接口回写到上述已注册和发布的 “天擎” 数据表, 并通过 MUSIC 接口查询读取数据,初步形成了数据、算力、算法在“天擎”环境下集约化管理和共享。客户端只需要通过接口直接读取预警数据并生成预警信息和发布,使得“端”的处理流程大大简化,程序代码大幅精简,实现了对“端”的轻量化改造。

4 结论

本文以实际业务需求为导向,遵循“天擎”业务融入标准和应用规范,基于“天擎”DPL 实现了对原有业务处理流程的改造融入, 一方面解决了实际业务中的需求,另一方面形成了一套可执行、可参考、实操性强的“天擎”DPL 设计与应用样例,为今后各气象业务系统的“云+端”融入改造积累了可借鉴的经验。

后续工作可基于“天镜”的 ES 库(ElasticSearch)自定义更多的 EI 和 DI 信息,以丰富“天镜”的监控内容; 同时积极开展基于 Docker 的 DPL 运行环境的研究,从而全面提升 DPL 的运行支撑环境的兼容

性，以利于后续更好地开展“天擎”融入工作。

“天擎”环境下的业务改造融入降低了对商业化软件框架的开发要求，支持多样化的开发语言并提供开发包，采用基于接口的开发形式降低了算法融入的开发难度。今后通过进行相关技术培训和不断深入探索，推动全区气象数据、算力、算法的集约化管理和高效共享，为形成“云上部署、终端应用”的气象云生态打下基础[11-17]。

参考文献：

[1] 国家气象信息中心.气象大数据云平台"天擎"1.0 应用开发手册(V1.2)[Z].2020.

[2] 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.

[3] 赵芳，熊安元，张小缨，等.全国综合气象信息共享平台架构设计技术特征[J].应用气象学报，2017，28(6)：750-758.

[4] 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.

[5] 赵冰燕，郭彩莲，来志云.基于青海气象大数据云平台的数裾服务接口[J]. 青海科技 ，2021，28(1)：82-86，90.

[6] 国家气象信息中心.气象大数据云平台“天擎”1.0 PORTAL 用户使用手册[Z].2020.

[7] 国家气象信息中心.MUSIC 管理接口开发使用手册[Z].2018.

[8] 国家气象信息中心.“天擎”资源清单-数据资源[Z].2020.

[9] 曾行吉，李涛，詹利群，等.基于 MUSIC 的特色数据与产品回写 CIMISS 方法研究[J].气象研究与应用，2018，39(1)：111-114.

[10] 赵芳，何文春，张小缨，等.全国综合气象信息共享平台建设[J].气象科技展，2018，8(1)：171-180.

[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，詹利群，任晓炜，等.Hadoop 环境下基于 SparkSQL 海量自动站数据查询统计的初探[J].气象科技，2019，47(5)：768-772，871.

[13] 李莉，廖剑伟，欧灵.云计算初探[J].计算机应用研究，2010，27(12)：4419-4422，4426.

[14] 刘凯悦.大数据综述[J].计算机科学与应用，2018，8(10)：1503-1509.

[15] 陈军成，丁治明，高需.大数据热点技术综述[J].北京工业大学学报，2017，43(3)：358-367.

[16] 黄志，李涛，宋瑶，等.基于 Json 的小型异构数据库同步策略研究[J].气象研究与应用，2020，41(1)：48-53.

[17] 黄志，丘平珠，谭斐. 区域自动站降水分时统计关键技术与实现[J].气象研究与应用，2019，40(2)：86-89.

Design and application of business integration based on CMADaaS DPL

Huang Zhi[1]，Huang Heng[2*]，Liang Weiliang[3]，Liang Yuanyuan[1]

(1.Guangxi Meteorological Information Center，Nanning 530022，China；2.Hechi Meteorological Bureau，Hechi Guangxi 547000，China；3.Guangxi Meteorological Observatory，Nanning 530022，China)

Abstract：According to the application integration standards and business specifications of CMADaaS，taking the processing process of meteorological short-term approaching heavy precipitation early warning service as an example，the business migration integration design process based on CMADaaS DPL processing line was given. The key links such as application integration account registration，public metadata registration，data storage design，DPL processing algorithm design and deployment，and DPL algorithm operation monitoring through "TianJing" were described，which provides technical ideas for the migration of meteorological business system to CMADaaS and the integration and transformation of "Cloud + Client".

Key words：CMADaaS；Cloud + Client；DPL；integration