

混合架构的可视化数据调度检索模型^①



孙周军^{1,2}, 乔文文¹, 侯 灵¹, 何文春², 张金标¹, 敖振浪¹

¹(广东省气象探测数据中心, 广州 510080)

²(国家气象信息中心, 北京 100081)

通讯作者: 敖振浪, E-mail: zhoujun_sun@tom.com

摘 要: 针对气象科研数据具有多源、异构、海量的特点, 中国气象局依托全国气象信息共享系统发布了统一数据访问接口服务为其提供数据支撑, 但实际使用中给科研人员获取数据也带来了技术难度. 通过采用混合架构研发的气象数据可视化调度检索系统, 将任务自动调度模型、可视化交互、数据格式适配和流量限制等技术相结合, 实现科研人员可视化、零编程的快速数据检索下载, 满足科研人员对数据格式的个性化需求. 实际应用表明, 该模型有效提升气象科研业务的数据组织效率, 协助了编程能力弱的用户对科研数据的使用, 调度模型设计思路和建立的数据共享业务流程在其他行业具有借鉴作用.

关键词: CIMISS; 服务接口; 任务模型; 计划与调度; 混合架构

引用格式: 孙周军, 乔文文, 侯灵, 何文春, 张金标, 敖振浪. 混合架构的可视化数据调度检索模型. 计算机系统应用, 2019, 28(12): 63-71. <http://www.c-s-a.org.cn/1003-3254/7202.html>

Visual Data Scheduling Retrieval Model Based on Hybrid Architecture

SUN Zhou-Jun^{1,2}, QIAO Wen-Wen¹, HOU Ling¹, HE Wen-Chun², ZHANG Jin-Biao¹, AO Zhen-Lang¹

¹(Guangdong Meteorological Observation Data Center, Guangzhou 510080, China)

²(National Meteorological Information Center, Beijing 100081, China)

Abstract: In response to the characteristics of multi-source, heterogeneous and massive data in meteorological research activities, the China Meteorological Administration has issued a unified data interface service based on China Integrated Meteorological Information Sharing System to provide data support for meteorological activities, there are certain technical barriers to data access in practice. In order to eliminate this technical barrier, the meteorological data visualization scheduling retrieval system developed based on hybrid architecture combines the task scheduling model, visual interaction and data format adaptation technology, and allows researchers to quickly access data in a visual and zero-programming manner. It meets individual needs for data formats. The application proves that the implementation of the model improves the data organization efficiency of the meteorological research business, assists the users with weak programming ability to use the data access interface, the ideas about the scheduling model and the business of data sharing can be used for reference in other fields.

Key words: CIMISS; aervice API; task model; plan and schedule; hybrid architecture

气象服务是国家公共服务和社会事业发展重要组成部分, 但气象的基础是数据. 气象服务已由传统的农业气象、地质灾害、水文气象、航空气象等不断向海洋、森林草原火险、交通、能源、城市、旅游、环

① 基金项目: 广东省气象局气象信息大数据创新团队 (201507); 广东省气象局科技项目 (GRMC2018M12)

Foundation item: Innovation Team for Meteorological Information Big Data of Meteorological Bureau, Guangdong Province (201507); Science and Technology Program of Meteorological Bureau, Guangdong Province (GRMC2018M12)

收稿时间: 2019-05-20; 修改时间: 2019-06-21; 采用时间: 2019-06-28; csa 在线出版时间: 2019-12-10

境、公共卫生等更广阔的领域拓展,科学家们从事的科学研究也向跨领域、跨行业发展^[1,2]。对专业的地球科学和社会其他行业数据采集、存储和管理,以及对互联网上海量的数据资源进行科学的超融合管理和便捷式供应成为气象科研活动需要迫切解决的问题。

为满足对海量气象数据管理和需求,中国气象局设计开发了全国气象信息共享系统(China Integrated Meteorological Information Sharing System, CIMISS),已于2016年底投入业务运行。该系统采用国、省两级布局,负责气象数据收集、加工处理、质量控制、存储管理和数据共享服务,为气象部门及相关行业用户快捷便利地获取气象数据提供数据支撑^[3-5]。全国各省份也在数据管理和需求方面做了相关研究,马渝勇等^[6]提出的气象信息管理与共享服务系统模型,基于中间件技术研发的接口服务让数据库的安全性和访问效率得到保障,实现了开放式的气象信息共享服务;王宏记等^[7]开发的气象信息综合数据库系统,从数据流程、采集、存储和质控等方面解决气象业务发展中面临的数据管理和需求问题;邓莉^[8]提出的基本气象资料共享系统,利用气象元数据管理技术,结合人机交互界面提供了可视化的资料下载共享服务;重庆、内蒙、河北等气象部门均研发不同的共享系统和数据服务。这些业务方法在一定程度上为气象数据的共享方法提供了改进的理论支持和实现方法,但未充分性考虑到科研用户的个性化调取和组织格式需求,未充分考虑需要较强的编程技能才能调取数据给科研人员带来的困惑等问题。CIMISS系统提供的数据接口平台提供了丰富的数据资源,提供多语言调用API方便开发人员调用数据,但科研人员需要较高的学习成本研究相关调用技术、以及对调取的数据进行二次处理形成所需的数据格式。有没有一种能够让科研人员零编程、自定义数据格式的调用机制,解决科研人员个性化需求呢?

本文提出了一种混合架构的可视化气象数据调度检索模型(IDEA Data Hub, IDH),该模型将可视化任务定制、调度和数据格式模板适配技术相结合,自动限流调取数据服务接口数据,较好的满足了科研人员对数据的个性化调用需求。

1 需求分析与设计

1.1 应用需求

气象科学研究是气象业务的发展基础,而在科研

究活动中,科学家一直受到科研数据组织的困扰,例如:不了详细的资料清单,不了解资料时间、空间、地域、观测频次和要素,不了解获得途径,不了解数据服务接口调用方法等。主要有以下原因:(1)数据虽爆发增长,但业务供应能力仍然偏弱。近年来随气象综合观测系统和互联网技术的迅速发展,来自于专业化观测和社会化萃取的气象数据,已经构成了气象科研活动所需的丰富数据资源。然而,依托CIMISS构建的国家和省级集约化数据环境,只是对传统的14大类进行了加工处理,且在线服务能力和存储容量有限,个别资料仅能提供十几天到几个月的在线服务。对气象预报、气候预测、决策服务和防灾减灾所需的跨行业、社会经济、社会化观测数据并未进行加工和管理。(2)数据科学管理,但与应用习惯脱节。同样依托CIMISS系统研发的统一数据访问服务对数据进行分类和元数据描述,解决了了解数据的途径问题,但输出的格式无法与科研工具对接,例如:MICAPS、Suffer、Grads、NCL、IDV等。(3)IT技术日新月异,科研人员跟进稍显滞后。计算机科学的3大定律能够很好的描述技术的发展速度,但气象科研人员更侧重于地球科学的原理、算法和实际应用的深入钻研,对于新的软硬件技术则关注较少,难免对于数据存储和服务采用的新技术和新方法无所适从。为解决以上问题,可通过计算机软件技术实现对数据下载、格式转换等流程的自动化封装,避免其对科研活动的干扰。

可视化气象数据调度检索模型的设计采用“云+端”的业务设计模式,确保科学家有丰富的数据资源取用,也保证数据检索下载的便捷性,解决科研人员组织科研数据的长期困惑。

1.2 总体设计

平台设计方案采用内嵌式Web服务的混合技术架构,分别保持C/S和B/S的架构的优势。表1详细列出2种架构的特点^[9,10],从中可见,C/S在图形的表现能力上以及运行的流畅度上远强于B/S模式,但B/S架构的跨平台、响应及时性、数据安全一致和实时性的优点很难替代。

未来发展趋势,是将两种架构优势完美地结合起来,该平台的应用系统既能以B/S的方式发布运行,同时又具有C/S方式的极强的可操作性,即混合架构,该架构恰好满足云+端系统的设计与实现需求。

考虑到气象科研人员对数据需求具有较大的个性

化需求,数据检索任务模型定义需要较强的交互操作,订阅任务、任务分类和执行记录的管理,以及订阅的数据需要本地存储等,所以需要采用 C/S 技术架构;同

时考虑到用户鉴权、数据清单浏览、接口元数据查看、过程数据清理配置等,具有通用性,所以采用 B/S 技术架构,方便统一研发和运维.

表 1 B/S 与 C/S 架构优缺点比较

架构	数据安全性	数据实时性	数据溯源性	服务响应及时性	存储模式	表现能力
C/S	节点分别存放数据,易发生数据丢失	多节点数据需要同步整合	不能随时随地看到业务发生状况	软件部署、升级等服务需要一定实施周期	后台数据库与本地存储结合使用,响应快	强
B/S	数据集中存放,比较安全	集中存放,不存在一致性问题	可实时查看当前业务发生所有变化	不可追溯	可快速更新升级	弱

1.3 功能设计

可视化气象数据调度检索模型以云+端模式业务运行,其中数据服务作为云端提供气象科研数据,可视化检索客户端软件作为终端提供数据下载调度的任务定制、管理、运行控制、格式转换和个性化设置等可视化交互操作.

1.3.1 数据服务接口

气象统一数据服务接口和实时历史一体化数据访问服务接口为气象数据调度检索模型提供气象数据服务,即数据服务云端.

(1) 气象数据统一服务接口 (Meteorological Unified Service Interface Community, MUSIC) 基于 CIMISS 面向气象业务和科研提供全国统一、标准、丰富的数据访问服务和应用编程接口,为国、省、地、县各级应用系统提供唯一权威的数据接入服务. 它的处理流程如图 1 所示. 用户在向 MUSIC 接口服务发起数据调用请求后,接口服务根据请求参数进行访问权限认证和数据访问权限认证,通过鉴权后,分别调用站点数据服务、格点数据服务和文件数据服务执行具体的数据获取与格式封装逻辑,从而原路返回给调用请求发起端.

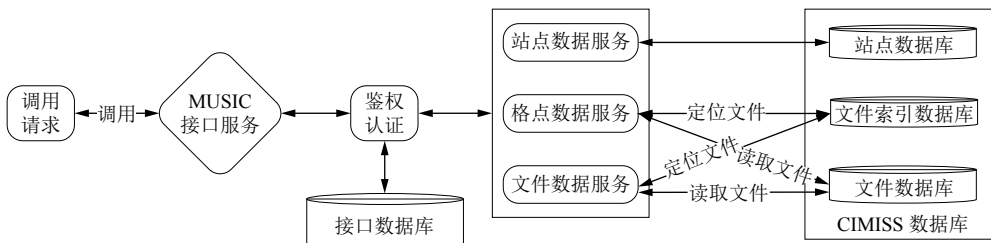


图 1 MUSIC 接口数据流程

(2) 实时历史一体化数据访问服务接口平台 (Integrated Database for Easy Access, IDEA) 数据服务接口,基于 CIMISS 系统标准自研的一体化集成数据环境 (包括结构化与非结构化气象数据),面向全省气象业务提供标准、统一的数据服务接口,是对 CIMISS 支撑服务能力的扩充. 它的处理流程如图 2 所示. 用户在向 IDEA 发起数据调用请求后,接口服务根据请求参数进行访问权限认证和数据访问权限认证,通过鉴权后,根据请求数据类别,将请求转发到具体的文件、格点、站点、图形数据服务和算法服务集群执行具体的数据获取与格式封装逻辑,从而原路返回给调用请求发起端.

从功能角度来说 IDEA 接口平台增加了检索路由功能、数据服务分布式服务功能,以及增加算法、图形产品数据服务,以及数据写入和删除 API 等功能;从具体的实现来说, MUSIC 是根据用户提交的参数动态拼接检索 SQL,而 IDEA 平台是预先定义的检索 SQL; CIMISS 数据库采用集中式存储,由于存储总容量限制则需定期对过期数据则进行清理,实时历史一体化数据库则采用分布式数据服务与分业务存储方式,理论上能够做到数据的永久存储和服务.

1.3.2 代理服务模块

气象数据可视化调度检索模型为数据服务接口平台智能化的调用者,设计并实现能够统一处理权限认证、接口元数据服务、调用请求转发、异常流量监控

其中 IDEA 接口与 MUSIC 接口平台相对比,从系

的代理服务,可在不介入接口服务平台的情况下实现对其兼容,即该模型需要对 MUSIC 和 IDEA 接口平台的兼容.具体处理流程如图 3 所示.

(1) 权限认证 模型自身并不提供用户权限认证,需要借助数据服务接口平台的账户认证功能确保气象数据的安全,所以在进行调度任务的创建时需要根据数据所属具体的服务接口平台进行鉴权,才能可见具有调用权限的数据接口列表.

(2) 元数据服务 数据服务接口平台的定义、属性和具体实现千差万别,统一定义该模型需要使用到的元数据结构体,并针对以上数据服务接口多态实现统一的元数据服务,在可视化调度检索模型上展示资料和接口列表,以及详细展示单个接口的调用系统参数和接口参数、以及其他说明信息.

(3) 请求转发 模型将具体的数据下载请求通过代理服务模块定位到具体的服务接口平台获取数据,在

转发请求的过程中增加额外的用于标识请求发起源的相关管理参数.

(4) 为避免随模型用户扩大和模型调度数据种类增长,对数据服务接口造成突发性访问压力,导致服务接口平台无法正常对用户提供服务,因此,模型在提交具体的数据请求时,按照预先设定的每秒请求量进行提交,当提交的请求量达到设定阈值时,则暂停提交请求.另外定时从服务接口平台的访问日志中分析时间段内处理的请求数量是否达到系统承受的压力,并提供可视化的强制用户禁用功能,优先保证数据接口服务稳定运行.

(5) 智能升级 模型最终是以客户端的方式运行在科研人员的终端主机上,软件需要具备能够监测自身是否具备更新条件、自动下载与重启,以及处理不同版本之间的兼容性等功能,智能升级模块需要对模型发出的以上请求进行相关的响应处理.

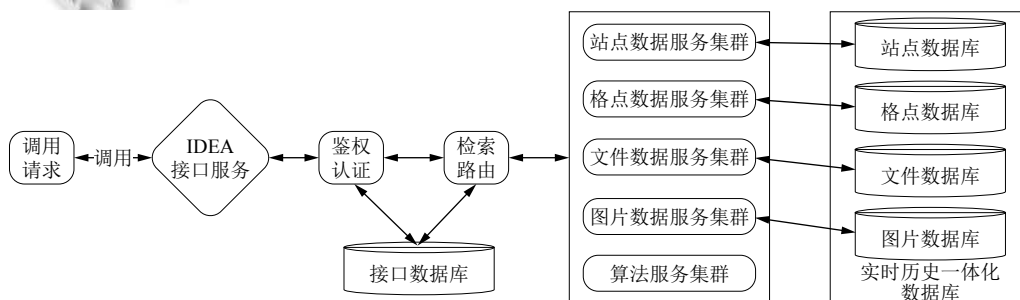


图 2 IDEA 接口数据流程

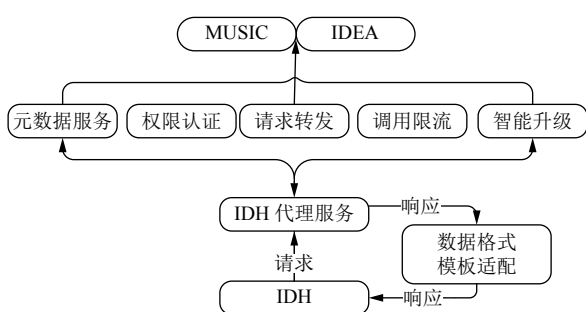


图 3 代理服务处理流程图

1.3.3 数据格式模板适配

数据格式模板适配功能是本模型的重要功能之一.在气象科研领域中常用数据格式有 Txt、Surfers、Micaps、NetCDF、HDF、Grib、BUFR 等,而气象数据服务接口输出的主要格式有 Txt、JSON、XML、Html、Csv 格式,要让用户直接获取到与应用切合度高的数

据,则规定只从服务接口获取 JSON 格式数据,并根据用户定制数据格式参数将其处理为应用可识别的数据格式,简化科研人员对数据的二次处理工作.转换功能基于插件方式进行具体实现,科研用户可自行实现所需格式的插件,进行功能扩展.

2 系统模型定义

气象可视化检索调度平台是 IT 技术在气象数据存储管理和应用中的科学综合运用,提供具有一定智能管理的可视化科学数据检索模型,由此需要定义任务的管理模型、任务状态迁移模型,以及设计科学的业务操作流程^[11,12].

2.1 任务模型定义

定义 1. 接口模型平台可适用于气象统一数据服务平台和广东的实时历史一体化数据访问服务平台,两

套接口平台的实现均采用 Restful 方式, 所以统一定义接口模型为 (Interface Model), 设接口对象属性组为 $I_n = \{I_nID, SP, IP\}$, 其中 I_nID 为接口唯一编号, SP 为接口调用的系统参数组, 包括账户、密码、数据返回格式, IP 为接口参数组, 根据具体接口的自身定义而定, 一般包括时次、地域、站点、经纬度范围等。

定义 2. 数据调用任务模型 (task model) 是科研人员定制的所有数据检索任务模型, 主要包括任务模型关联的接口模型和任务调度的相关属性等, $TM = \{T_mID, I_nID, \{sT, eT\}, tmS, \{eI, eD\}, rT, \{pE, fP, bF\}\}$. 其中 T_mID 为任务模型编号, sT 、 eT 为模型有效时间定义, 其中 eT 可以取值为空, 则表示任务模型永久生效, tmS 为模型的运行状态, eI 为任务调度的间隔, eD 为任务调度时匹配时间参数的偏移量, rT 为重试次数, pE 、 fP 为任务模型存储数据的路径和文件名表达式, bF 为任务模型下载的数据是否合并为大文件. 任务模型分为历史数据调取模型、未来数据调取模型和混合数据调取模型, 主要取决于任务模型的有效时间区间。

定义 3. 数据调用任务实例 (task instance) 是任务模型产生要执行的具体任务, 包括任务模型的具体执行时间点、执行状态、已重试次数、任务实例具体下载数据的时间范围, 以及任务模型所有属性的赋值, $TI = \{T_nID, exeT, exeS, exeF, rT, \{pE, fE, bF\}, \{pST, pET\}\}$, 其中 T_nID 为任务实例编号, $exeS$ 为任务执行状态, $exeFT$ 为任务执行失败次数、 pST 与 pET 分别为检索数据的时间范围。

定义 4. 数据调用任务实例集 (task instance set) 是所有任务实例集合. 平台根据科研人员定义的任务模型生成的需要执行的任务实例列表, 分为立即执行实例集和未来执行实例集. 其中立即执行实例集用来调取创建任务模型时间点前数据的任务实例, $TISI = \{TI\}$, 未来执行实例集用于调取创建任务模型时间节点后数据的任务实例, $TISW = \{TI\}$, 则任务实例集定义为: $TIS = \{TISI, TISW\}$.

2.2 任务模型中的角色

任务模型管理器 (Task Model Manager, TMM), 负责任务模型运行过程中的任务状态转换控制, 根据任务模型定义生成具体的任务实例集, 监视统计任务实例集已完成、已生成和总数, 同时为便于科研人员对任务模型的管理, 提供创建、删除、修改、克隆、启动/暂停和手工生成任务实例的功能. 其中任务模型状态转换如图 4 所示。

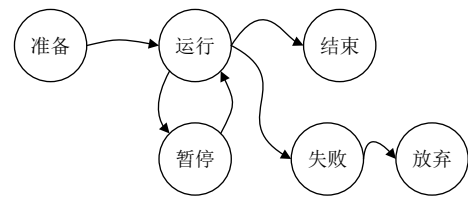


图 4 任务模型状态转换

任务实例管理器 (Task Instance Manager, TIM), 负责收集需要立即执行的任务实例集, 任务实例的具体执行 (包括: 调用接口获取数据、将返回结果按照任务模型定义组织文件存储等)、状态转换和失败重试, 以及避免并发过多给服务端带来较大压力, 所以 TIM 负责任务实例调度的并发流量控制等功能。

任务调度管理器 (Task Schedule Manager, TSM), 主要负责对平台中的已完成任务实例、带执行任务实例、失败任务实例和暂停任务实例进行定期清除; 负责对该平台向接口服务发送调用请求量的统计, 便于服务端能够对请求量大的用户进行强制禁用, 确保接口服务能够正常运行。

2.3 任务管理流程

为了更直观、更详细的展现任务模型调度的实现过程, 将该过程用流程图展示出来, 如图 5 所示。

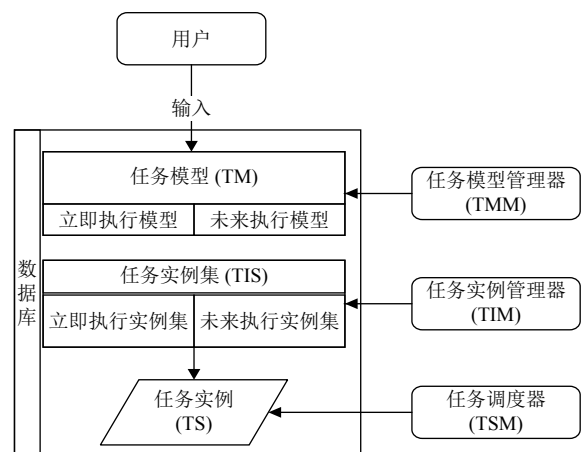


图 5 任务管理流程图

(1) 任务实例生成过程 用户根据选定的接口, 设定任务模型需要填写的调度参数和接口参数, 任务模型管理器 TMM 保存任务模型. 任务实例管理器 TIM 根据用户定义的任务模型立即生成任务实例集, 对于未来任务实例集只生成当天实例, 后续的未来任务实例则由任务实例管理器定期预先生成。

(2) 任务实例调度过程 任务实例管理器定时线程

对所有实例集按照预执行时间倒序排序,按照预设的流量阈值检索一定数量的任务实例,设定为 C ,通过下列规则将任务实例提交到任务调度器 TSM, TSM 则根据任务实例的预定执行时间执行具体线程. 设 t 为 TSM 定时检索任务实例间隔; i 为批量提交任务间隔; v 为设定的流量阈值,为 i 时间内提交的任务数量; 则 $c = v * (t/i)$ 为每次处理最大数量. 如果 $C = c$,则系统满负荷提交任务; 如果 $C < c$,则依据距预执行时间权重阶梯性确定提交任务量.

(3) 任务实例重试过程 任务实例执行后,根据执行结果状态进行后续处理. 如果任务执行失败,则执行次数自动加 1,且将该任务实例的预定义执行时间延长一个 t 值,等待 TSM 的下次处理. 如果执行次数属性值达到系统设定的放弃执行阈值,则将该任务实例设定为放弃状态不再进行处理.

3 关键技术

3.1 任务实例集预处理实现

根据任务模型定义,每日定时生成第二天需要执行的任务实例列表. 任务模型定义中的属性包括: 任务有效时间范围、任务调度间隔、任务调度时匹配时间参数偏移量,计算与执行的任务时间和接口的具体时间参数. 具体的任务实例生成算法如下:

算法 1. 任务实例生成算法

Task instance generate procedure(), 此过程中的参数表示 t 时刻任务模型集.

1. 取得系统时间 t ;
2. $S = \{TM_0, TM_1, \dots, TM_n\}$;
3. $n = |S|$;
4. For $i=0$ to n
5. $TM = S_i$;
6. If ($t \geq TM.sDate$ && $t < TM.eDate$ && $TM.exeS! = FINISH$), 则执行步骤 7, 否则, 执行步骤 4;
7. 初始化任务实例 TI 对象;
8. 计算任务实例执行时间;
9. 计算任务实例下载数据的时间范围;
10. 初始化任务实例状态和已重试次数;
11. 根据 $TM.exeT$ 和 $TM.Pe$ 计算存储路径;
12. 根据 $TM.fp$ 、 $TM.bF$ 和 $exeT$ 计算存储文件名;
13. 保存任务实例项 TI ;
14. 循环执行步骤 4.

3.2 任务实例执行限流实现

漏桶算法是网络世界中流量整形或速率限制时经常使用的一种算法,主要是控制提交到接口服务的调

用请求流量,避免突发流量对对服务接口造成类 DDoS 攻击. 为此设计了任务实例调度限流算法,具体步骤如下:

算法 2. 任务实例调度限流算法

1. lastTime = 取得最先需要执行的任务实例时间;
2. waitExeTaskSet = 获取 lastTime 时间偏移 N 的任务实例集;
3. Executor.submit(waitExeTaskSet);
4. Sleep(T);
5. 循环执行步骤 1.

3.3 数据压缩优化传输效率

MUSIC 和 IDEA 接口采用 HTTP 协议提供文本数据服务,整个过程中网络传输占据较大的时间. 因此,数据调度检索模型每次调用数据时,强制在客户端与服务器之间采用压缩方式传输数据. 经实验测试,采用标准的 GZIP 方法可压缩至原内容的 40%,从而节省 60% 的数据传输. 任务实例在向服务接口发送数据调用请求和处理数据的具体过程如下:

- (1) 初始化请求, IDEARequest request = new IDEARequest();
- (2) 设定请求采用 GZIP 算法传输数据, request.setHeader("content-type", "gzip");
- (3) 提交请求, 获取经过压缩传输的数据, gzipReturn = request.post();
- (4) 对 Gzip 压缩数据进行解压缩处理, unGzipReturn = unGzipDecoder(gzipReturn).

3.4 数据格式转换实现

MUSIC 和 IDEA 接口服务提供标准的 JSON 格式数据结果,要对其进行转换处理为 MICAPS、Suffer、Grads、NCL、IDV 等软件所需的格式. 基于工厂模式实现的数据格式转换插件化模块,具体实现过程如下:

- (1) 定义格式转换接口协议 DataFormat, 并定义 json2DataFormat (gzipReturn) 方法;
- (2) 实现各个数据格式转换插件, 插件必须实现 DataFormat 协议;
- (3) 实现格式转换插件工厂, 创建具体的插件实例;
- (4) 按照 3.3 节方法下载具体数据并解压;
- (5) 格式转换工厂类加载具体需要转换格式类, 将解压数据作为转换类的输入参数, 完成格式的转换. 具体方法如下:

```
DataFormat df = DataFormatFactory.getInstance()
.createDataFormat("MICAPS"); // 加载 Micaps 数
```

据格式转换插件

```
String micapsReturn=df.json2DataFormat
(unGzipReturn); //对解压 JSON 数据进行格式转换
```

3.5 交互性设计

零编程获取气象数据是该模型设计的目标之一,所以设计功能完整、操作便捷的可视化交互界非常重要.按照用户的使用习惯和软件设计方法,如图6所示,系统分为用户登陆、接口浏览、任务创建、任务管理、个性设置、智能升级和强制下线7个部分.

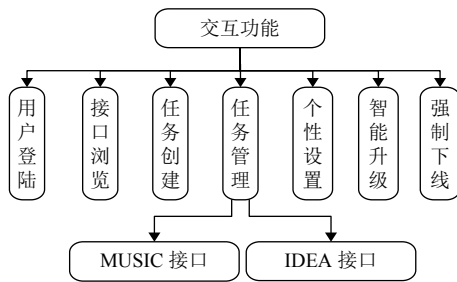


图6 气象数据调度检索软件功能

(1) 用户登陆,即用户在使用该模型时,首先要获得 MUSIC 或 IDEA 接口平台的接口调用账户,并在该模型进行登陆后方可操作其他功能,同时可以进行两套账户的登陆切换和注销.

(2) 用户登陆,即用户在使用该模型时,首先要获得 MUSIC 或 IDEA 接口平台的接口调用账户,并在该模型进行登陆后方可操作其他功能,同时可以进行两套账户的登陆切换和注销.

(3) 接口浏览,可以通过代理服务获取到接口服务平台发布的所有分类和接口列表,查看接口的详细定义:系统参数和接口参数等.

(4) 任务创建,用户在选定调用数据接口后,点击“创建固定间隔任务计划”或“创建定时任务计划”,则通过定制任务模型,并选定需要保存的文件格式.

(5) 任务管理,为方便科研用户进行任务模型和实例的管理,模型按照任务计划、已完成任务、待执行任务、失败任务和暂停任务进行分组,用户可对任务进行业务分类,用户可精确查看任务的执行情况.

(6) 强制下线,为保护数据服务接口对模型发起的调用请求量,除过进行自身流量控制之外,还定时通过调用日志进行分析,对于近1小时内的用户进行访问量和数据下载量的监视,管理员认为可疑时,则可强制该模型下线,从而控制其不在对数据服务接口发起访

问请求,需要用户与管理员沟通重新开通.

(7) 个性设置,因为任务模型的定义、任务实例、任务执行结果等信息均记录在模型的内嵌数据库中,用户可根据实际情况设定具体清除策略,以及开发新的数据格式转换插件进行部署.

4 性能分析与业务应用

4.1 性能测试

从中国气象局 CIMISS 系统调取全国 50 000 多气象自动站整点小时数据,组织为 MICPAS 格式为例进行测试,分别调取 1、3、6、12 个月的数据进行耗时和性能测试,结果如表2所示.

表2 检索数据性能测试

时间段	网络流量 (Mb)		耗时 (分)		访问次数
	压缩	未压缩	压缩	未压缩	
2018.01-01	23	46	4	13	744
2018.01-03	35	179	31	90	2232
2018.01-06	80	269	58	180	4464
2018.01-12	278	540	120	400	8928

根据测试结果,气象数据调度检索客户端采用压缩传输时,从总体耗时和数据传输量均有显著提升.其中组织 1 年的数据仅需 2 小时,业务上完全可以满足科研活动需求.同时,限流策略的应用,以及对服务接口的并发监视,并未对其造成突发性的访问的压力.

4.2 运行效果

最终实现的软件运行效果如下所示.图7是用户登陆后能够浏览到已经定义的所有任务模型和实例,以及接口服务提供的所有资料接口信息.



图7 任务模型和实例显示界面

图8是用户定制获取 2018 年中国地面逐小时文本格式的调度任务.

创建固定间隔任务计划

调度参数

任务计划名称: getSurfEleByTimeRange 任务分组类型: 非共享任务

数据检索起始日期: 2018-01-01 00:00:00 数据检索结束日期: 2019-01-01

重试次数: 4 任务延迟执行时间(单位: 分钟): 5

文件保存路径: D:/IDEADData/getSurfEleByI
在路径最后添加日期格式, 可使文件按照日期保存示例: D:/IDEADData/yyyy/MM/dd

文件名: Micaps_ 保存为单个文件

查询间隔时间(单位: 分钟): 5 时间参数加减(单位: 小时): 0

接口参数

URL: http://10.148.89.55/cimiss-web/api?interfaceId=getSurfEleByTimeRange&userId=

dataFormat: text

图8 任务定义示例界面

图9 是用户查看所有任务模型的运行状态列表, 以及对任务模型进行控制的功能界面。

计划编号	计划名称	执行状态	开始时间	结束时间	创建时间	操作
55	海事局台风灾害数据	未启动	2018-11-21 00:00:00	2018-11-28 23:59:59	2018-11-28 17:32:17.754	详情 暂停 开始 取消 删除 手工启动任务 详细
56	海事局台风灾害数据	已启动	2018-11-20 00:00:00	2018-11-20 23:59:59	2018-11-20 07:01:23.454	详情 暂停 开始 取消 删除 手工启动任务 详细
57	海事局台风预报数据	未启动	2018-11-21 00:00:00	2018-11-20 07:01:11.879		详情 暂停 开始 取消 删除 手工启动任务 详细
58	海事局台风预报数据	未启动	2018-11-20 00:00:00	2018-11-20 23:59:59	2018-11-20 07:01:03.746	详情 暂停 开始 取消 删除 手工启动任务 详细
51	电力局气象数据	未启动	2018-11-21 00:00:00		2018-11-20 06:55:55.879	详情 暂停 开始 取消 删除 手工启动任务 详细

图9 任务模型控制界面

4.3 典型应用

目前, 气象数据可视化调度检索模型软件已经在多个业务中进行应用, 均产生好的应用效果. 行业用户如民航、军队、水利、农业等, 以及政府部门对气象观测数据的需求非常旺盛, 每年数据共享量达到 100 TB, 但气象部门近年来综合观测系统的迅速发展, 观测的原始数据格式逐步与向世界气象组织标准靠拢, 每次数据格式的改革均给资料使用单位带来重新解码处理等诸多不便, 该模型作为生成共享格式的文件中间环节, 避免了观测资料格式变化带来的共享问题, 具体业务流程如图 10. 目前, 已经承担向 20 多个单位提供标准的共享数据文件. 且已经在海洋预报数值模式、南

海台风预报、珠三角洲大气灰霾等科学研究领域, 以及海珠城市生态气象综合系统^[13]等多个建设项目中使用本模型组织获取科学数据, 满足业务需求.

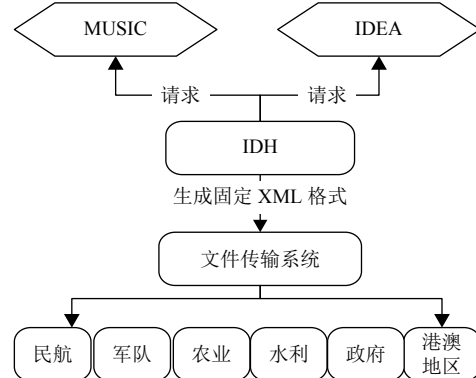


图10 基于 IDH 数据共享业务流程

5 结论与展望

可视化气象数据调度检索模型针对气象科研人员的获取科研数据的痛点问题, 结合可视化交互、任务调度和数据格式模板适配技术, 为科研提供便捷、零编程的数据组织方法, 提高科研活动前期数据准备的效率.

经过在气象业务与科研活动的实际应用, 对于全国气象部门以 MUSIC 和 IDEA 接口提供数据支撑的业务模式起到很大的推动作用; 在向非气象部门进行数据共享中, 屏蔽数据源结构的异构性和变化, 避免用户因数据格式或存储结构变更导致业务变动的困扰, 提升气象行业的数据服务信息化水平. 同时, 任务调度模型设计可用于其他领域的业务调度模型中; 对于形成的云+端数据交换流程, 适用于其他行业领域的数据共享业务中, 为充分发挥数据的价值提供一种新的共享方法.

参考文献

- 黄瑞芳, 周园春, 鞠永茂, 等. 气象与大数据. 北京: 科学出版社, 2017.
- 沈文海. 云时代下的气象信息化与管理. 北京: 电子工业出版社, 2017.
- 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. [doi: 10.11898/1001-7313.20170610]
- 赵芳, 何文春, 张小纓, 等. 全国综合气象信息共享平台建设. 气象科技进展, 2018, 8(1): 171-180. [doi: 10.3969/j.issn.2095-1973.2018.01.023]

- 5 徐拥军, 何文春, 倪学磊, 等. 基于 MUSIC 接口的 CIMISS 数据环境故障应急切换设计与实现. 气象科技进展, 2018, 8(1): 213–218, 231. [doi: 10.3969/j.issn.2095-1973.2018.01.041]
- 6 马渝勇, 徐晓莉, 宋智, 等. 省级气象信息共享系统的设计与实现. 应用气象学报, 2011, 22(4): 505–512. [doi: 10.3969/j.issn.1001-7313.2011.04.014]
- 7 邓莉, 王国复, 孙超, 等. 基本气象资料共享系统建设. 应用气象学报, 2004, 15(S1): 33–38.
- 8 王宏记, 王海军, 曾又枝, 等. 省级气象信息综合数据库系统的设计与实现. 暴雨灾害, 2008, 27(3): 283–286. [doi: 10.3969/j.issn.1004-9045.2008.03.016]
- 9 范霖君. 基于 C/S 和 B/S 混合模式信息系统的研究与设计. 电子技术与软件工程, 2015, (15): 204–205.
- 10 曹晟, 蔡自兴. 基于 C/S 与 B/S 混合软件体系结构的封闭式管理系统的设计. 计算机工程与应用, 2004, 40(5): 224–226. [doi: 10.3321/j.issn:1002-8331.2004.05.072]
- 11 邓子云, 王如龙, 杨晓峰, 等. 基于 SOA 集成平台的 EFSM 任务调度模型. 计算机工程, 2011, 37(23): 238–240. [doi: 10.3969/j.issn.1000-3428.2011.23.081]
- 12 刘应明, 廉东本. 数据交换平台的任务调度模型设计. 小型微型计算机系统, 2013, 34(7): 1543–1547. [doi: 10.3969/j.issn.1000-1220.2013.07.018]
- 13 王刚, 孙周军, 吴映, 等. 基于 IDEA 数据平台的海珠城市生态气象综合系统建设. 广东气象, 2016, 38(2): 77–80. [doi: 10.3969/j.issn.1007-6190.2016.02.019]