高分遥感共性产品生产系统关键设计

张正，李宏益，胡昌苗，唐娉

中国科学院空天信息创新研究院, 北京 100094

摘 要：以众多遥感卫星的发射与行业应用系统为代表，遥感应用技术在近年间取得了快速发展，国产高分系列卫星等平台所提供的数据越来越丰富，以遥感共性产品作为信息源的应用系统也已经覆盖了多个行业。然而这两者之间的桥梁，即由国产卫星数据生产标准化系列遥感共性产品的能力，仍显得相对薄弱，极大限制了国产遥感数据的使用率与影响力。这一能力的改善需要从产品体系、产品算法与生产系统等多方面共同推进。本文从高分遥感共性产品生产系统的角度，针对系统所面对的来自数据集成、算法集成、生产编排与云计算等方面的诸多挑战，提出了一套完整的系统关键设计。系统在容器化集成运行的基础上，涵盖了集群、数据、算法、权限、参数、工作流、生产等一系列关键环节的关键技术。系统研发与生产实践表明，本系统可以流畅的集成各类共性产品算法，实现稳定高效的业务化生产平台，助力于国产卫星产品服务能力的有效提高。

关键词：生产系统，共性产品，算法集成，容器，云计算，系统架构，工作流

中图分类号：P2

引用格式：张正,李宏益,胡昌苗,唐娉.2023.高分遥感共性产品生产系统关键设计.遥感学报,27(3): 651-664

Zhang Z, Li H Y, Hu C M and Tang P. 2023. GF quantitative remote sensing production system: Core design. National Remote Sensing Bulletin, 27(3): 651-664[DOI: 10.11834/jrs.20210420]

1 引 言

遥感应用技术近年来在中国取得了快速发展，这种发展尤其以居于上游的卫星平台和占据下游的各行业应用为典型，二者分别作为遥感应用的入口与出口，相关工作形成了较强的影响力与显示度。在卫星平台领域，由于在过去很长时期内遥感数据一直处于缺乏状态，卫星发射与数据接收系统一直都是优先发展的目标，目前国内已经拥有了GF（白照广，2013；潘腾，2015）、ZY（Zhang等，2014）、HJ（Wang等，2010）、FY（Yang等，2009）等系列卫星，观测手段逐渐丰富，时空分辨率不断提高，特别是在10—30 m尺度，国产卫星数据的丰富程度已经优于国外。在业务化应用领域，由于遥感观测具有范围广、周期短、受限少等优点，遥感共性产品作为数据基础已经被广泛应用于公共卫生、能源、环境监测、防灾减灾、农业、生态、气象等多个行业（Deng等，2019；Hu等，2019，2020；Shao等，2019），针对行业自身业务需求，众多遥感业务化应用系统也已经被建立。随着高分数据不断普及，高分遥感处理与应用的发展也十分迅速（Kang等，2018；聂娟 等，2018；Zhong等，2021；周建民等，2021）。

然而处于中游的国产遥感卫星产品体系与生产系统，在一定程度上成为了薄弱环节，目前海量的国产卫星数据尚未能有效转化为规范、实用、连续、稳定的遥感共性系列产品，极大限制了国产卫星数据的使用率和影响力，行业用户仍无法摆脱对国外成熟遥感共性产品的依赖，或分散了大量财力精力自研共性产品及其生产系统。只有形成国产卫星遥感共性产品体系，建立稳定业务化运行的大规模遥感共性产品生产系统，才能架起国产卫星数据与行业应用之间的桥梁，最大化卫星数据的使用效益，服务于国家重大战略领域需求，支撑区域与全球共同发展。

收稿日期：2020-09-25；**预印本：**2021-07-19

基金项目：国家重点研发计划(编号:2018YFA0605500)；国家自然科学基金青年科学基金(编号:41701399)；高分辨率对地观测系统重大专项(编号:21-Y20B02-9003-19/22)

第一作者简介：张正,研究方向为遥感图像处理、大数据挖掘、高性能计算。E-mail：zhangzheng@aircas.ac.cn

通信作者简介：唐娉,研究方向为遥感图像处理、人工智能与深度学习。E-mail：tangping@aircas.ac.cn

共性产品是指介于标准产品与专题产品之间，承上启下的定量化应用产品，主要包括（1）标准产品经几何精纠正、正射校正得到的数据应用产品，例如数字表面模型、数字正射影像等；（2）反映地表反射、辐射、散射特征的应用产品或用于大气校正的大气参量产品，例如地表反射率、气溶胶光学厚度等；（3）反映地球能量、植被、水和大气特征的数据应用产品，例如地表温度、叶面积指数、水体透明度和大气二氧化碳浓度等。

高分遥感共性产品（柳钦火 等，2023）生产系统由于遥感数据与遥感算法自身的多样性与复杂性等特点，以及近几年间云计算软硬件基础环境的持续跨越式演进，在技术上引发了诸多新的挑战，具体包括如下方面：

算法集成：遥感共性产品算法具有较强的专业性，多样性与复杂性，这些特点体现在算法实现中就会造成不同算法使用的开发语言、函数库、辅助软件、运行环境等方面具有较大差别，亟需一种能使算法在不同软硬件环境下无差别运行的算法集成运行框架（张正 等，2016；Li和Tang，2020）。

数据集成：不同卫星数据除了数据内容本身的差异外，在数据格式、投影方式、文件组织、元数据描述等数据构成形式上也有较大差别，从系统平台的角度，还需要考虑数据可视化等其他应用场景，因此需要能够同时面向定量计算与可视化的多源数据一体化存储组织与管理技术（Lü等，2011；Ma等，2015；Li等，2020）。

生产编排：遥感共性产品生产算法具有计算时间长、数据吞吐量大、输入数据种类繁多等特点，而生产工作流则具有层级化、批量化、数据依赖等特点，需针对上述特点抽象出一套不同算法参数的统一描述方法，以及复杂嵌套工作流的调度引擎，共同支撑稳定高效的业务化生产运行（Yan等，2018）。

云计算：目前云计算技术已经成熟落地，随之而来的是云虚拟化主机、云对象存储、容器、微服务架构等一系列新软硬件环境，遥感共性产品的生产对于算力与存储一直都有较高要求，生产系统需要及时进行技术革新，向云计算架构迁移，借助云计算技术进一步提升生产效能（Wang等，2018）。

目前成熟度较高的遥感共性产品生产平台多是面向特定卫星传感器的，例如生产MODIS产品的EOSDIS系统（Earth Observing System Data and Information System）（Esfandiari等，2007），以及Landsat（Irons等，2012；Roy等，2014）、Sentinel（Drusch等，2012）产品相应的生产系统，这些系统的设计较少考虑多源融合产品的生产，无法同时利用分辨率相似的不同传感器数据。多源协同定量遥感产品生产系统MuSyQ（Multi-source Synergized Quantitative Remote Sensing Production System）（张正 等，2016；柳钦火 等，2018）在全球、中国—东盟、“一带一路”及重点试验区4个不同尺度上集成了遥感数据的归一化处理与20余种多源融合产品流程，但受限于当时的技术发展，其系统设计没有充分利用最新的服务化、容器化与云计算技术，系统各环节都面临技术升级与改造。在云计算技术逐渐普及的过程中，一些基于云的多数据中心遥感定量产品生产系统被陆续提出（Wang等，2018；Fan等，2018），利用分布式集群实现了高效的海量数据管理与大规模任务调度，但同样由于当时的数据资源与技术限制，上述系统主要面向MODIS等中低分辨率卫星数据，且算法主要部署在虚拟机上，没有采用更轻量的容器技术。随着高分数据源的不断丰富与容器云等新技术的快速发展，如何充分利用新一代高性能计算技术实现高分共性产品大规模快速业务化生产成为了亟待解决的问题。

为了应对上述各方面对于遥感共性产品生产系统提出的新要求与新挑战，本文提出了一套完整的系统关键设计。系统围绕容器技术（Bernstein，2014）实现遥感算法与服务的统一集成运行，容器技术是云计算中的核心技术，可以实现高效轻量级的进程隔离，通过将算法相关的所有运行环境与依赖项打包成镜像，赋予算法容器广泛的可移植性与一致性。在容器化运行的基础上，本文对系统相应的功能组成与架构、容器化集群、算法镜像仓库、算法权限、算法封装与集成、算法参数描述、数据组织、共性产品生产、工作流调度等核心环节的关键技术都进行了设计，以共同打造一个稳定高效可集成各类算法的运行化生产系统，有效提升高分遥感共性产品生产服务能力。

2 系统架构与关键设计

2.1 系统功能组成与架构

本系统的总体功能架构分为四层，如图1所示，自底向上分别为基础资源层、核心功能层、服务接口层与界面交互层。

基础资源层为系统提供基础软硬件支撑，提供网站的运行环境与算法的计算环境，支持在各类云平台上的部署，同时也兼容传统的分布式集群部署。基础硬件环境主要包括云主机与云存储、基础软件环境主要包括容器平台Docker（Merkel，2014）、容器编排管理系统Kubernetes（Burns等，2016）、关系数据库MySQL、微服务平台Spring Cloud、以及算法镜像仓库Harbor。

图1 系统总体功能架构图

Fig. 1 Functional architecture of the proposed system

核心功能层包括系统核心的3大子系统、数据引接管理子系统、运行管理子系统以及产品生产子系统。数据引接管理子系统包含数据与产品的入库、编目、剖分、发布、查询、可视化等全生命周期的管理功能，对多源、异构、不同尺度、不同投影的空间数据进行统一存储组织和管理，建立与生产体系相适应的时空数据快速索引结构，为产品生产提供快速的数据检索、存储、获取和一致性校验服务。运行管理子系统包含对系统运行所涉及的各类业务对象的综合管理功能，包括对用户、权限、配置、订单、任务、系统状态、节点、算法以及工作流等的管理，对不同算法与工作流的参数和结构进行统一描述，建立用户与算法权限分级体系，构建算法镜像仓库，制定生产订单提交标准流程，实现节点与系统状态的实时反馈，为系统运行提供的全方位的管理支撑。产品生产子系统包含与算法容器化运行以及生产订单执行相关的全链路功能，包括对生产任务的规划，生产工作流的解析、分层与调度，算法各组件的容器化封装，算法镜像的推送与拉取，Kubernetes任务配置清单的自动生成与提交，基于Kubernetes的容器任务编排与调度，任务执行状态的实时监控等，共同实现容器化计算集群之上的遥感共性产品稳定高效生产。

服务接口层将系统功能组织成对外提供服务的接口，主要包括负责算法与工作流上传、编辑、展示的算法与工作流服务接口、负责订单创建与产品生产的订单生产服务接口、负责地理时空可视化的网络地图服务接口、负责用户信息与权限管理的用户服务接口、以及负责数据与产品管理发布的数据与产品服务接口。

界面交互层为不同角色的用户提供相应的网页交互界面、主要包括算法上传页面组、算法展示页面组、工作流定制页面组、订单创建页面组、订单查询页面组、数据入库页面组、数据检索页面组、以及用户设置页面组等。

2.2 容器化计算集群

本系统中的遥感算法均以容器化方式运行，相应的在底层由一到多个容器化计算集群提供软硬件支撑。遥感共性产品算法具有显著的数据存储量大，数据读写任务繁重、计算时间长、算法种类多样等特点，本系统在设计集群架构时充分考虑了上述特点，提出了一种能够灵活部署于各类云上与线下环境的容器化集群结构，集群软件均采用云原生技术，简洁务实并且保持了与其他外在系统或技术最大化的开放性与兼容性。图2展示了一个集群的典型组成架构，主要包括四类组成部分，分别为主控节点、计算节点、算法仓库与数据存储。

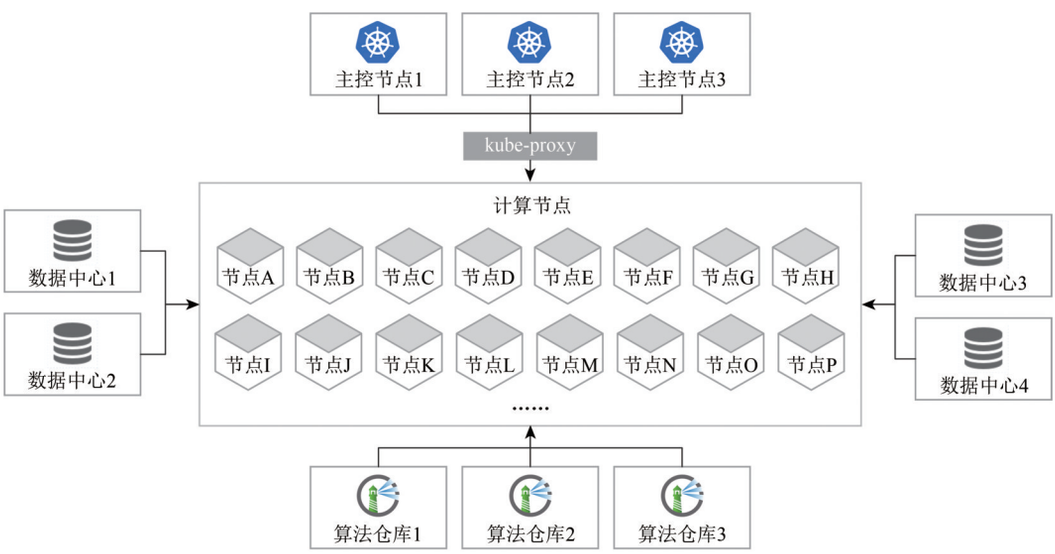


图2 容器化集群架构

Fig. 2 Architecture of containerized clusters in the system

主控节点使用Kubernetes容器编排系统实现对算法容器的管理调度，为了防止主控节点失效，提高系统可用性，同时分流业务压力，实现负载均衡，每个集群配备3个并行的主控节点。

计算节点使用原生的Docker容器平台运行算法容器，并通过kube-proxy网络服务与主控节点保持通信，系统支持计算节点数量的动态弹性扩展，以应对遥感算法长时间的计算高峰。

算法仓库基于Harbor镜像仓库构建，提供即时的算法镜像拉取与推送服务，采用https协议以保障网络传输安全，为了实现高可用性与负载均衡，以应对遥感算法的多样性，集群并行部署3个互为备份的算法仓库同时提供服务。

数据存储也采用多中心的部署方式，以分流繁重的数据读写任务，系统支持多种异构的本地、在线或云存储形式，与算法镜像仓库的不同之处在于数据存储不完全是系统内部私有的，而经常是接入性质的，直接接入外部提供的数据存储，因此多个数据存储中心提供的数据也不是互为镜像的，系统内部控制的数据存储会考虑数据备份与缓存，以分流对同一份数据短期内的集中请求。

主控节点、计算节点、以及算法仓库节点均对底层设施没有特别要求，可以构建在一般物理机、虚拟机或云主机之上，或者采用混合架构，以实现灵活部署与扩容。

2.3 算法镜像仓库

本系统作为遥感共性产品算法的集成平台，算法镜像仓库是其核心组件之一。本系统采用Harbor构建系统内部私有的算法镜像仓库。如图3所示，系统内的算法按照功能可以分为产品算法、处理算法、工具算法和评价算法4类，每一类算法镜像分别保存在相应的子仓库中。

产品算法即直接生产遥感共性产品的算法，保存在/product子仓库中，主要包括几何产品算法、辐射基础产品算法、土地覆盖产品算法、植被参量产品算法、能量平衡参量产品算法、水分参量产品算法、大气参数产品算法等；

处理算法即与遥感数据处理相关的算法，保存在/processing子仓库中，主要包括多源多尺度的几何归一化算法、辐射归一化算法、大气校正算法、云检测以及云修补算法；

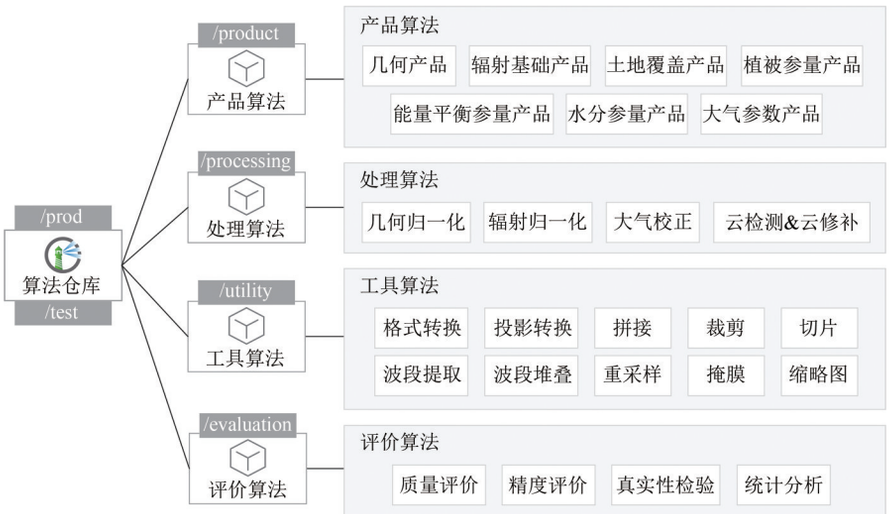


图3 算法镜像仓库架构

Fig. 3 Architecture of algorithm image repositories in the system

工具算法即具有工具性质的遥感数据或产品操作算法、保存在/utility子仓库中，主要包括格式转换、投影转换、拼接、裁剪、切片、波段提取、波段堆叠、重采样、掩膜、缩略图等操作所对应的算法，这些算法根据所适用的数据产品类型可以进一步细分；

评价算法即对数据或产品进行分析验证的算法，保存在/evaluation子仓库中，主要包括质量评价算法、精度评价算法、真实性检验算法、以及统计分析算法等，这些算法可根据所要验证的数据或产品种类进一步细分。

算法镜像仓库在测试环境（test）和生产环境（prod）中分别独立部署，在测试环境中成熟的算法可以迁移至生产环境。为了实现高可用性与负载均衡，算法镜像仓库在测试和生产环境下分别设置了3个副本，彼此保持同步。算法镜像仓库的访问采用https安全传输协议，仓库将本地签署的数字证书发布到需要访问仓库的节点，以授权访问。

算法镜像仓库中算法的命名模式约定为｛算法库域名地址｝/｛子仓库名称｝/｛算法名称｝:｛版本标签｝，例如harbor.prod-1.com/product/ndvi:v1.0即表示在域名为harbor.prod-1.com的生产环境镜像仓库下的产品算法子仓库中版本为v1.0的NDVI算法。在整个系统中的任何场景下，例如生产任务脚本的编写或算法封装推送，均可以使用如上的算法名称唯一指名算法。

2.4 算法权限

遥感共性产品算法是研究人员知识的结晶，对算法知识产权的界定与保护应该成为集成生产系统设计中的核心关切。本系统基于用户角色对算法的权限进行了明确划分，并且算法权限的决定权完全属于算法的归属用户。系统对于算法的操作主要分为查看与使用两种，查看即可以通过算法的展示页面查看算法的详细信息，知晓该算法；而使用即可以将算法应用于真实的生产过程，生产共性产品。

算法归属用户可以选择的算法权限有四种，如表1所示，分别为开放、授权、私有以及不公开。开放权限表示算法对于所有用户都是可见且可用的，是开放程度最高的权限；授权权限表示算法对所有用户可见，但只对指定的用户可用，例如归属用户可以赋予同实验室或同单位的用户算法使用权，而限制其他用户对算法的使用；私有权限表示算法对所有用户可见，但只有算法的归属用户可以使用该算法进行产品生产；而不公开权限则表示算法是完全封闭的，算法对其他用户均不可见，只有算法的拥有者可以使用该算法进行产品生产。上述算法权限可以通过例如Shiro等用户安全框架进行实现。

表1 算法权限设置

Table 1 Algorithm authorization settings

算法权限/操作	查看	使用
开放	所有用户	所有用户
授权	所有用户	指定用户
私有	所有用户	归属用户
不公开	归属用户	归属用户

2.5 算法封装与集成

本系统中的算法以容器镜像的形式进行集成，以实现跨语言跨环境的一致性运行，算法集成的业务逻辑如图4所示。

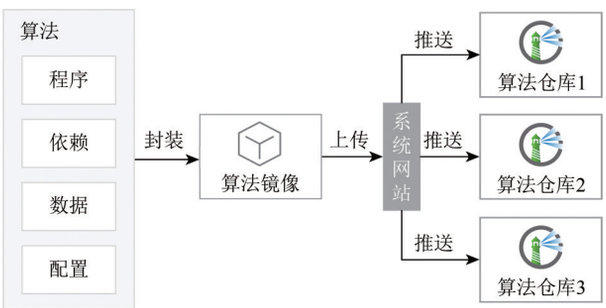


图4 算法封装与集成的业务逻辑

Fig. 4 Business logic of algorithm encapsulation and integration

首先将算法运行所需的各组件打包封装成算法镜像，对于遥感算法，常见的组件主要包括可执行程序、所依赖的函数库或第三方软件、算法私有的辅助数据、配置文件或环境变量等，可以包含多个可执行程序，但需指定一个主程序作为入口，第三方函数库或软件需设置引用路径，使其能够被镜像内的程序所调用，辅助数据如果过大，可以考虑作为算法的输入参数由镜像外引入，避免算法镜像文件整体过大造成传输障碍。

系统约定的算法镜像结构如图5所示，即安排了算法各组件在镜像内所存放的位置，算法镜像在根目录下创建/app文件夹作为算法的根目录，其中又包括lib、data、bin、config共4个子文件夹，分别部署算法的依赖库、辅助数据、可执行程序以及配置项。存放依赖库的lib路径一般需要被添加到系统的$LD_LIBRARY_PATH环境变量中，以使依赖库能够被系统发现；存放可执行程序的bin路径需要被添加到系统的$PATH环境变量中，以使可执行程序也能够被系统发现。

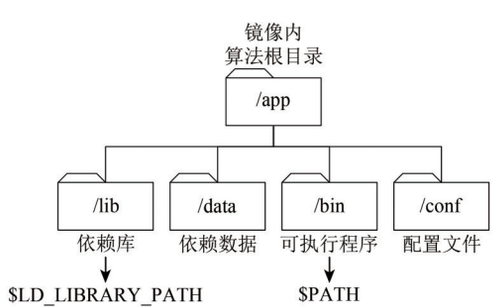


图5 算法镜像内的文件组织结构

Fig. 5 File structure inside algorithm docker images

在算法各组件都齐备的情况下，将算法打包封装成镜像的过程需要通过Dockerfile来定义，Dockerfile是一种描述镜像构建过程的文本文件，其内容包含了一系列构建镜像所需的设置和指令，构建镜像时主要的步骤包括指定基础镜像、创建文件夹、将组件拷贝至镜像内部的指定位置、设置环境变量与权限、设置入口程序及默认参数等。为了编写的便利，系统按照约定的算法镜像结构，提供基础版本的Dockerfile，包括了镜像文件结构的创建和与文件架构相应环境变量的设置，后续还需要根据算法实际情况将各组件拷贝到约定位置，并完成算法入口设置等必要的操作。而后通过docker容器平台的docker build命令运行编写完整的Dockerfile即可实现算法镜像的打包封装。

算法镜像封装完成后，通过docker save命令将镜像以文件的形式保存，即可通过网页前端将镜像上传至系统，系统自动将镜像推送至算法镜像仓库，并在多个仓库间进行同步，达到可用状态。在上传算法时除了算法镜像，还需要提交算法相关的元信息，例如算法简介、作者、单位、样例图、数据格式、算法输入输出参数描述等。

2.6 算法参数描述

本系统支持用户提交的算法参与生产流程，新提交的算法可与其他算法连接以形成工作流，一种算法耦合程度最轻的连接方式是通过输入输出数据进行连接，即之后算法的输入是之前算法的输出。为了使算法能够顺畅的与系统内的其他算法或数据资源进行衔接，系统需要定义一套对所有算法统一的输入输出参数描述方式，算法在提交时只需要按照这一方式描述其输入输出参数，并且算法输入输出数据的格式已知，系统即可使该算法与其他算法进行连接。

遥感算法的参数一般有两类，分别为值参数（value parameter）与数据实体参数（data entity parameter）。值参数直接取参数的字面值，例如数字或字符串等类型的参数，无需更多描述；而数据实体参数表示算法对一类数据实体的需求，在算法调用时系统需要根据算法对数据的具体需求从数据库中检索匹配的数据实体，并将数据实体的存储地址作为参数的值。

基于对多种遥感算法输入输出参数的抽象，

本系统采用如表2中所述的九个属性描述一个数据实体参数。

表2 描述数据实体参数的9个属性

Table 2 Nine properties for data entity parameter description

参数名称	说明
数据类型	此参数代表的数据类型，例如MCD12Q1
参数序号	此参数在算法所有参数中的位置序号
是否外来产品	此类数据是否只能由外部导入，抑或可以被系统生产，如可以生产，则会在找不到某个数据时尝试触发生产流程
空间分辨率	此类数据的空间分辨率
时间分辨率	此类数据的时间分辨率
时间跨度	需要多长时间跨度内的此类数据
时间对齐方式	以算法所要生产产品的时间为中心，往哪个方向查找数据，有三种选择，向前查找、向后查找、或向两侧查找
是否分幅	需要的是否是分幅后的数据
网格类型	分幅所采用的网格类型

在描述算法输入参数时以上9个属性都是必要的，而在描述输出参数时，“是否外来产品”、“时间跨度”、“时间对齐方式”这3个属性则不再被需要，因为能够被算法生产的产品必定不是外来产品，而输出数据不需要在数据库中进行检索，因此就无需知道其检索时的时间跨度和时间对齐方式。在给出了数据实体参数的完整描述后，系统即可以在数据库中按照算法的需求检索输入数据。系统将算法所有参数的描述组织在一个XML文件中，上传算法时随算法其他组件一并上传。

2.7 数据组织

作为全球遥感共性产品生产系统，不可避免的会面临海量数据、多种投影方式、多种分辨率尺度、多时相和多文件格式等数据存储组织上的难点，另一个问题是，当前面向数据可视化和定量处理的数据组织基础架构不一致，如果对海量数据在显示和处理时分别采用不同的组织方式，则会带来显著的冗余存储和性能损耗，因此本系统提出一套同时面向显示和定量处理等多应用场景的多源遥感数据一体化存储组织技术，该技术通过构建多尺度遥感数据剖分体系，高效的数据缓存策略和基于OGC（Open Geospatial Consortium）扩展的数据访问接口，使同一套遥感数据即可以面向可视化又可以面向定量化处理，并能将投影转换的次数控制在1次以内，且能屏蔽多源数据格式的复杂性。

具体的，为提高数据的访问效率，系统依据数据类型、数据时间、访问层数等设计缓存数据的存储结构，缓存数据采用轻量级的本地Sqlite数据库存储，缓存瓦片数据的存储和访问地址都是由数据类型、时间、层、全球统一框架的地理位置行列数计算唯一Hash值。Hash算法是一种散列算法，可以将计算出的数值等概率均匀分布到不同区间，从而减少访问冲突，提高存储空间利用率，提高查询效率。此处设置Hash值长度为32位，前24位用于确定Sqlite数据库文件，后8位用于定位Sqlite文件中的瓦片文件，这样的Hash分层分片存储技术可以有效避免碎小文件，加速缓存寻址效率，提高数据服务性能。

系统通过拓展的OGC数据接口提供同时面向多应用场景的数据服务。传统OGC接口都是面向数据可视化的，一般只提供红绿蓝三波段的瓦片数据，而定量处理请求数据的波段不限，且波段顺序也不定，甚至需要请求不同传感器的数据，因此系统设计上下两个层次的数据接口，下层接口与OGC服务模式一致，上层接口将下层OGC接口返回的数据聚合成定量处理直接可以使用的数据，我们称之为RTU（Ready To Use）数据接口。上下两层接口设计既可以满足数据可视化的需求，同时又可以满足遥感定量处理的需求。

由于多尺度、不同投影间数据的转换缺乏统一基准，数据处理过程中经常需要使用多次投影转换，导致数据精度损失，因此系统同时设计实体数据的剖分规则和与可视化的对应关系，实体数据剖分时，与现有遥感数据的投影、剖分方式尽可能保持一致，将数据协同应用和显示过程中的投影转换都控制在1次以内，并将可视化时的不同层级的响应数据与数据实体建立对应关系。

2.8 共性产品生产

遥感共性产品生产是本系统的核心业务场景，本系统采用订单驱动的生产模式，具体的业务流程如图6所示。用户通过创建订单页面首先选择所要生产的产品种类，对于每一种产品，都可能会有多种生产算法，分别由不同的团队研发、上传并维护，每种算法也会有不同版本，因此在选择产品种类后还需要选择具体算法种类的具体版本。

如具体的算法版本涉及可定制的参数值，例如某个可调节的阈值，则在指定具体算法版本时在界面上也可以自主设置这些参数的值，当然系统也提供默认的参数值。

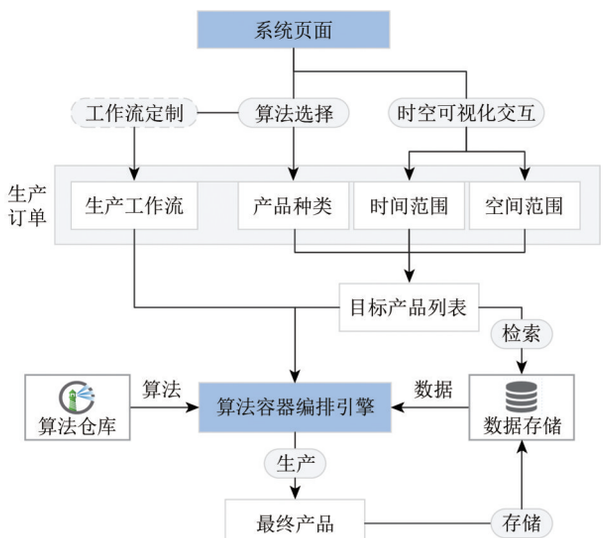


图6 订单提交与产品生产业务逻辑

Fig. 6 Business logic of order submission and production

系统内许多产品的生产依赖其他产品作为输入，且产品间的依赖关系还可能嵌套多层，例如植被净初级生产力（NPP）的生产依赖光合有效辐射分量（FPAR），而FPAR的生产又依赖叶面积指数（LAI）和光合有效辐射（PAR），LAI和PAR的生产又依赖标准产品，前述的完整生产流程即组成了生产工作流。本系统支持用户定制从标准产品开始的生产工作流，具体的，支持用户选择最终产品算法，对于算法的各项产品依赖分别再指定产品算法，如此递推直至只存在对于标准产品的依赖，此时即完成了用户指定的各种算法之间的连接，组成了产品工作流。对于工作流中的各算法也可以自主调节算法开放给用户的参数。

如果用户只关心直接生产目标产品的算法而不关心其输入数据的生产，则可以不定义完整的工作流，系统会使用所有同种类的产品并采用预设的生产算法。

在明确了目标产品及其生产算法或工作流之后，用户还需要选定所需产品覆盖的时间范围和空间范围，系统在前端页面提供时空可视化交互的方式支持用户选择时间范围与空间范围。至此用户完成了订单四要素的确定：（1）产品种类，即生产何种产品；（2）生产算法或工作流，即如何生产该种产品；（3）时间范围；（4）空间范围。

每种产品的时间分辨率和空间分辨率都是明确的，在确定了订单产品种类，时间范围与空间范围之后，对于标准分幅的产品即可确定目标产品列表，列表具体到每一景需要生产的产品，因此在产品实际生产之前，系统即可将产品列表反馈给用户，提高系统友善度与透明度。

系统接收产生订单后会开始产品生产，生产过程中每个算法在执行之前，会根据其对输入数据的需求在数据库中检索输入数据并确认数据文件的完整性，如输入数据缺失，则会尝试通过系统自行进行生产，如系统自身无法生产，在外部接口条件具备的情况下会尝试引接外部数据以尽可能满足生产条件。算法对输入数据的确认发生在每个算法执行之前，作为算法运行的一部分，而不是通过额外的数据规划阶段，这样可以使每个算法的运行更加独立，与算法调度功能的边界更加清晰，使系统能用一种解耦且统一的方式处理各种算法，而不会造成各种特例，打乱系统的一致性。

算法容器在计算集群中的编排调度采用Kubernetes实现，Kubernetes是一套容器化应用的自动部署、编排和管理系统，它通过一系列不同控制器，实现容器的不同执行策略，本系统中的每个算法对应一个Kubernetes中的Job控制器，Kubernetes通过一个YAML格式的任务配置清单描述所要执行的算法任务，清单中需要对任务使用的算法镜像、计算资源、存储资源、网络资源等进行描述，本系统会自动生成任务配置清单，并提交给Kubernetes执行。

每个具体的可执行算法镜像在算法仓库中都具有全局唯一标识，通过此标识可以从算法仓库中实时拉取算法镜像而后运行，由于本系统采用Kubernetes进行容器编排，因此只需在配置清单中指定算法标识，算法拉取会在算法调度后由负责算法运行的节点自动进行。产品生产完成后会将元信息与数据文件保存至数据库与数据存储中。

2.9 工作流调度

遥感共性产品生产具有批量化的特点，一般都是对某一区域范围下达生产任务；同时产品生产工作流本身具有层级化的特点，高层级产品以低层级产品作为输入。考虑到上述批量化与层级化的特点，本系统提出一种分层集约的工作流批

量调度策略，即将同一批次的多个工作流合并考虑，根据工作流执行的逻辑顺序，将每个串行阶段中可以被并行执行的所有算法归为一层，层中的所有算法任务均执行完毕后再进入下一个串行阶段，直至最后一个串行阶段。

图7展示了NPP生产工作流的分层集约方式，由于FPAR的生产依赖LAI和PAR，而这两者的生产没有其他依赖，因此生产LAI和PAR成为了第一个串行阶段，所有工作流中生产LAI和PAR的算法实例归为第一层，层内算法之间可以并行执行，第一层中所有的算法实例完成之后，可以进行FPAR的生产；类似的，接下来NPP的生产依赖FPAR，故所有生产FPAR的算法实例归为第二层；最后所有生产NPP的算法实例归为第三层，等第二层中的所有算法实例完成之后，再启动第三层的生产。

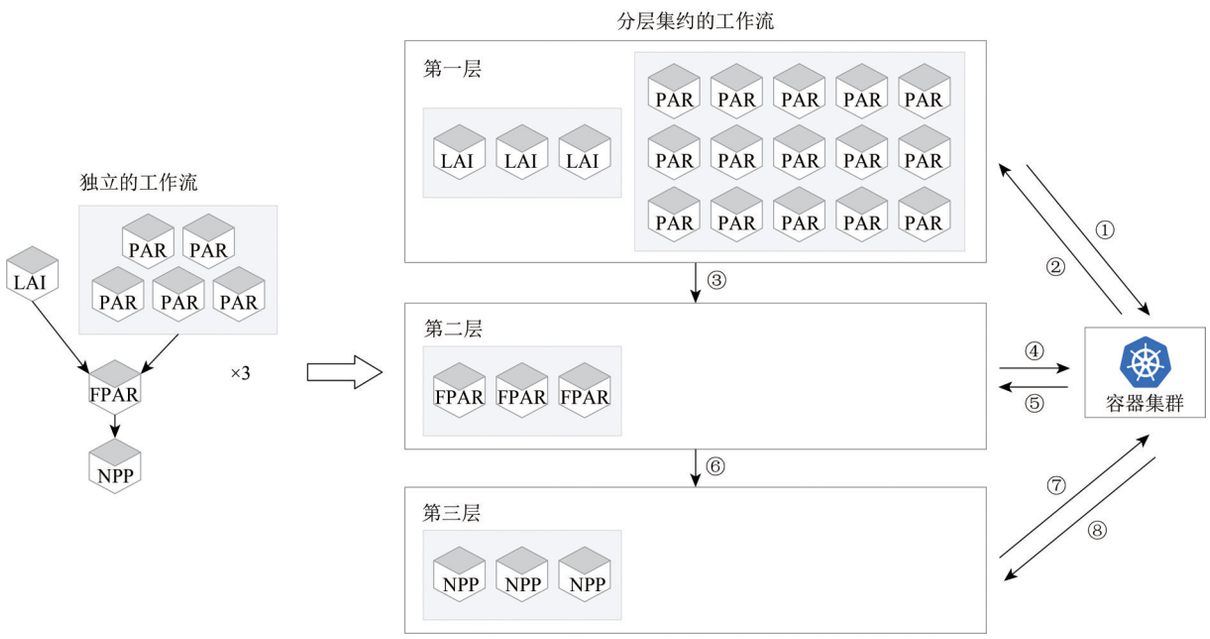


图7　批量工作流的分层集约调度策略（①②等数字代表执行步骤的顺序号）

Fig. 7　The stratified aggregation dispatching strategy for workflow batches (numbers like ①② represent the sequential order of execution)

分层集约策略将多个工作流中的算法进行统一分层、统一调度，以批次为单位进行整体性的生产，可以有效降低多工作流调度的复杂度，减轻系统资源消耗，同时层内算法依旧保持了高并行度，在维持算法并行度的基础上进一步提高了生产效率，相关的数据与计算资源也以批次为单元进行调度，从整体的角度优化了系统生产流程。

3　系统实例与产品生产

系统的主要功能为大时空尺度下的高分遥感影像处理与共性产品生产，本章通过对系统与产品生产实例的介绍，验证本文所提出关键技术的应用效果与能力。

3.1　高分共性产品算法集成

依照本文设计实现的生产系统目前已经集成了8类典型高分共性产品算法，并利用高分1号WFV宽幅卫星数据完成了2013年—2020年连续8年中国区域的产品生产。

表3展示了系统当前支持的高分共性产品列表，包括地表反射率标准产品（REF）、植被指数产品（NDVI）、叶面积指数产品（LAI）、植被覆盖度产品（FVC）、光合有效辐射吸收比例（FPAR）、植被总初级生产力（GPP）、地表反照率（Albedo）以及土地覆盖分类（LC）。上述算法目前主要采用高分1号WFV宽幅卫星数据作为输入，因此空间分辨率均为16 m，时间分辨率除了包含单景数据外，主要采用10 d合成的方式，土地覆盖分类产品每季度进行更新。当前产品最高层级为3级，地表反射率是经过几何与大气校正后的1级标准产品，植被指数、叶面积指数、光合有效辐射吸收比例、地表反照率、土地覆盖分类以地

表反射率作为输入，因此是2级产品，植被覆盖度与植被总初级生产力需要叶面积指数作为输入，因此是3级产品。后续系统还将集成更多产品类型，并支持算法版本持续更新。

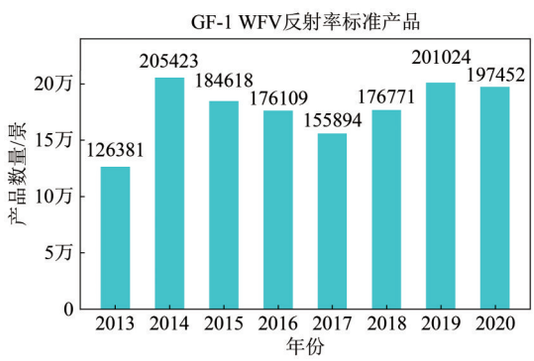
表3 系统当前高分共性产品列表

Table 3 GF product list of the current system

产品名称	空间分辨率/m	时间分辨率
地表反射率	16	单景
植被指数	16	单景/10天
叶面积指数	16	单景/10天
植被覆盖度	16	单景/10天
光合有效辐射吸收比例	16	单景/10天
植被总初级生产力	16	单景/10天
地表反照率	16	单景/10天
土地覆盖分类	16	季度

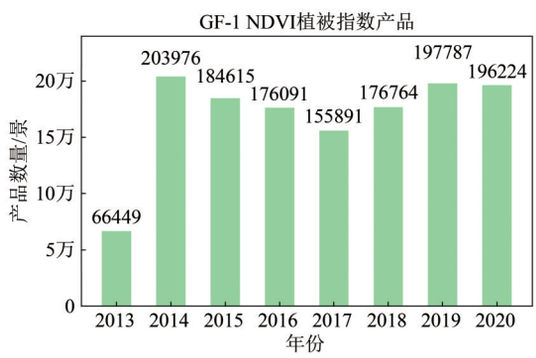
3.2 中国区域2013—2020高分共性产品生产

截止本文成稿时，系统已完成对中国区域2013—2020连续8年地表反射率、植被指数、叶面积指数、地表反照率4类高分共性产品的生产。图8（a）—（d）分别展示了4类产品的逐年单景数量统计，产品均采用UTM投影下的1°×1°网格进行分幅，此处计数为分幅后的产品数量。经统计，系统共生产了1423672景地表反射率标准产品，总存储量约为200 TB，共生产了1357797景植被指数产品，总存储量约为23 TB，共生产了1190709景叶面积指数产品，总存储量约为15 TB，共生产了1385394景地表反照率产品，总存储量约为50 TB。



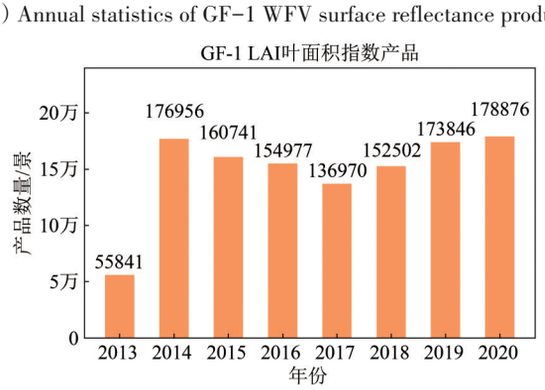
（a）GF-1 WFV 地表反射率产品逐年数量统计

（a）Annual statistics of GF-1 WFV surface reflectance product



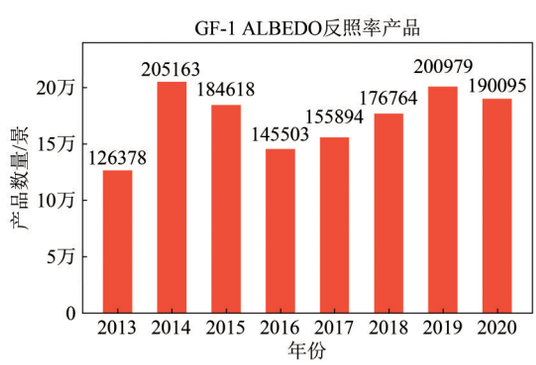
（b）GF-1 WFV 植被指数产品逐年数量统计

（b）Annual statistics of GF-1 WFV vegetation index product



（c）GF-1 WFV叶面积指数产品逐年数量统计

（c）Annual statistics of GF-1 WFV leaf area index product



（d）GF-1 WFV 地表反照率产品逐年数量统计

（d）Annual statistics of GF-1 WFV surface albedo product

图8 系统产品逐年数量统计

Fig. 8 Annual statistics for number of products of the system

图9展示了2015年—2020年单景高分植被指数产品的数量—空间分布，即每年每个空间网格内的单景产品数量，越接近红色表示产品数量越多，越接近紫色表示产品数量越少，可以观察到系统产品每年均对中国区域进行了完备覆盖，大多数网格在一年间都有一两百次的产品覆盖。

百万景TB级的高分产品生产实践表明本文提出的关键技术与系统具备较强的生产与处理能力，能够支持高分遥感共性产品的生产需求。

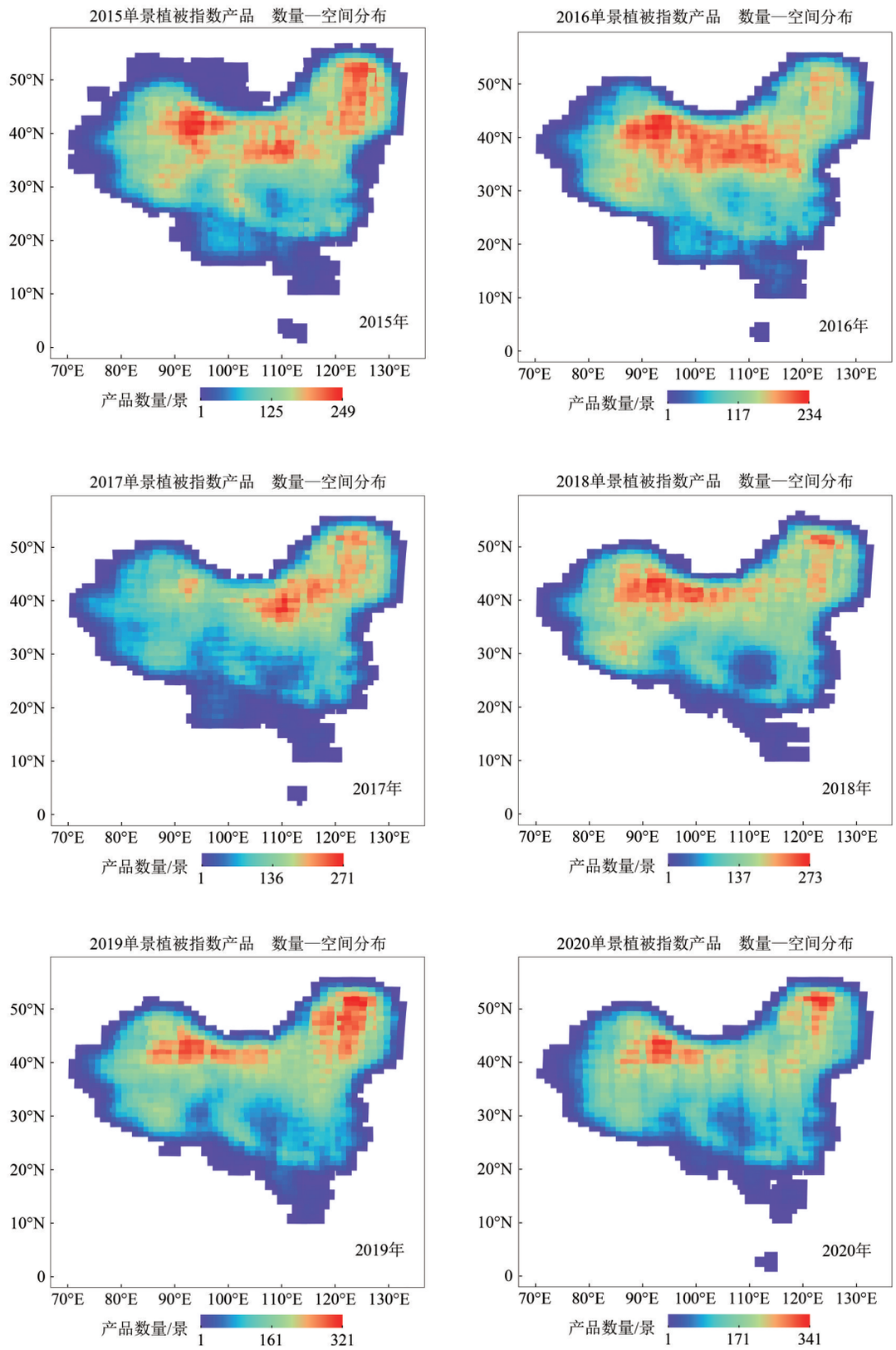


图9 中国区域2015年—2020年单景植被指数产品数量—空间分布

Fig. 9 Quantity and spatial distributions of single-scene vegetation index products from year 2015 to 2020 in China area

4 结 论

本文提出了一套高分遥感共性产品生产系统关键设计，根据遥感数据与算法的特点，在系统各相应环节均做出了针对性处理。对于遥感算法实现与部署的多样性，系统采用容器化的方式对算法进行封装并实现跨语言统一集成运行；为了支持大规模容器化计算，系统建立了多个容器镜像仓库与多中心的容器集群，同时引接多个数据中心的数据，实现数据源覆盖与负载均衡；为保护算法知识产权，系统设置了用户可灵活选择的基于角色的算法权限；为使用户上传算法直接接入系统并检索系统中的输入数据，系统提炼了统一描述数据实体参数的属性列表；同时面向数据可视化与定量计算，系统提出了多场景的数据统一组织管理策略；对于生产过程中常见的批量工作流，系统给出了分层集约的工作流批量调度优化策略。系统持续支持了多种大尺度数据处理与产品生产实践，各项系统设计在实践中均发挥了预期效果，表明基于本系统设计可以实现稳定高效支持云生态的新一代运行化生产系统，助力于高分遥感产品服务能力的实质提升。

参考文献（References）

Bai Z G. 2013. Technical characteristics of the GaoFen-1 satellite. Aerospace China, (8): 5-9 (白照广. 2013. 高分一号卫星的技术特点. 中国航天, (8): 5-9)

Bernstein D. 2014. Containers and cloud: from LXC to docker to kubernetes. IEEE Cloud Computing, 1(3): 81-84 [DOI: 10.1109/MCC.2014.51]

Burns B, Grant B, Oppenheimer D, Brewer E and Wilkes J. 2016. Borg, omega, and kubernetes: lessons learned from three container-management systems over a decade. Queue, 14(1): 70-93 [DOI: 10.1145/2898442.2898444]

Deng X D, Liu P H, Liu X P, Wang R Y, Zhang Y Y, He J and Yao Y. 2019. Geospatial big data: new paradigm of remote sensing applications. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(10): 3841-3851 [DOI: 10.1109/JSTARS.2019.2944952]

Drusch M, Del Bello U, Carlier S, Colin O, Fernandez V, Gascon F, Hoersch B, Isola C, Laberinti P, Martimort P, Meygret A, Spoto F, Sy O, Marchese F and Bargellini P. 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote sensing of Environment, 120: 25-36 [DOI: 10.1016/j.rse.2011.11.026]

Esfandiari M, Ramapriyan H, Behnke J and Sofinowski E. 2007. Earth observing system (EOS) data and information system (EOSDIS)—evolution update and future//2007 IEEE International Geoscience and Remote Sensing Symposium. Barcelona: IEEE: 4005-4008 [DOI: 10.1109/IGARSS.2007.4423727]

Fan J Q, Yan J N, Ma Y and Wang L Z. 2018. Big data integration in remote sensing across a distributed metadata-based spatial infrastructure. Remote Sensing, 10(1): 7 [DOI: 10.3390/rs10010007]

Hu T, Renzullo L J, Cao B, Van Dijk A I J M, Du Y M, Li H, Cheng J, Xu Z H, Zhou J and Liu Q H. 2019. Directional variation in surface emissivity inferred from the MYD21 product and its influence on estimated surface upwelling longwave radiation. Remote Sensing of Environment, 228: 45-60 [DOI: 10.1016/j.rse.2019.04.012]

Hu T, Renzullo L J, Van Dijk A I J M, He J, Tian S Y, Xu Z H, Zhou J, Liu T J and Liu Q H. 2020. Monitoring agricultural drought in Australia using MTSAT-2 land surface temperature retrievals. Remote Sensing of Environment, 236: 111419 [DOI: 10.1016/j.rse.2019.111419]

Irons J R, Dwyer J L and Barsi J A. 2012. The next Landsat satellite: the Landsat data continuity mission. Remote Sensing of Environment, 122: 11-21 [DOI: 10.1016/j.rse.2011.08.026]

Kang W C, Xiang Y M, Wang F, Wan L and You H J. 2018. Flood detection in Gaofen-3 SAR images via fully convolutional networks. Sensors, 18(9): 2915 [DOI: 10.3390/s18092915]

Li H Y and Tang P. 2020. Dps-MuSyQ: a distributed parallel processing system for multi-source data synergized quantitative remote sensing products producing. IEEE Access, 8: 79510-79520 [DOI: 10.1109/ACCESS.2020.2989138]

Li H Y, Zhang Z and Tang P. 2020. A web-based remote sensing data processing and production system with the unified integration of multi-disciplinary data and models. IEEE Access, 8: 162961-162972 [DOI: 10.1109/ACCESS.2020.3021791]

Liu Q H, Wen J G, Zhou X, Zhao J, Li Z Y, Li X, Ma M G, Wang W Z, Liao X H, Liu S M, Fan W J, Xiao Q, Zhong B, Li J, Xin X Z, Li L, Jia L, Gao Z H, Jin J D, Liang S, Xin J, Liao C J and Wu Y R. 2023. Technique system of remote sensing product generation and validation of GF common products. National Remote Sensing Bulletin, 27(3): 544-562 (柳钦火, 闻建光, 周翔, 赵坚, 李增元, 李新, 马明国, 王维真, 廖小罕, 刘绍民, 范闻捷, 肖青, 仲波, 李静, 辛晓洲, 李丽, 贾立, 高志海, 金家栋, 梁师, 邢进, 廖楚江, 吴一戎. 2023. 高分遥感共性产品生成和真实性检验技术体系. 遥感学报, 27(3): 544-562) [DOI：10.11834/jrs.20235022]

Liu Q H, Zhong B, Tang P, Zhang H H, Li H Y, Wu S L, Xin X Z, Li J, Jia L, Shan X J, Zhang Z, Wen J G, Du Y M, Li L, Yang A X, Li H, Hu G C, Zhao J, Zhang H L, Yu S S, Dou B C and Wu J J. 2018. Remote sensing data products oriented quantitative computing

system—the GSC best practice data computing environment 2018. Journal of Global Change Data and Discovery, 2(3): 271-278 (柳钦火, 仲波, 唐娉, 张宏海, 李宏益, 吴善龙, 辛晓洲, 李静, 贾立, 单小军, 张正, 闻建光, 杜永明, 李丽, 杨爱霞, 历华, 胡光成, 赵静, 张海龙, 余珊珊, 窦宝成, 吴俊君 . 2018. 多源协同定量遥感产品生产系统——2018年中国地理学会地理大数据计算环境“优秀实用案例”. 全球变化数据学报, 2(3): 271-278) [DOI: 10.3974/geodp.2018.03.04]

Lü X F, Cheng C Q, Gong J Y and Guan L. 2011. Review of data storage and management technologies for massive remote sensing data. Science China Technological Sciences, 54(12): 3220-3232 [DOI: 10.1007/s11431-011-4549-z]

Ma Y, Wu H P, Wang L Z, Huang B, Ranjan R, Zomaya A and Jie W. 2015. Remote sensing big data computing: challenges and opportunities. Future Generation Computer Systems, 51: 47-60 [DOI: 10.1016/j.future.2014.10.029]

Merkel D. 2014. Docker: lightweight Linux containers for consistent development and deployment. Linux Journal, 2014(239): 2

Nie J, Deng L, Hao X L, Liu M and He Y. 2018. Application of GF-4 satellite in drought remote sensing monitoring: a case study of Southeastern Inner Mongolia. Journal of Remote Sensing, 22(3): 400-407 (聂娟, 邓磊, 郝向磊, 刘明, 贺英 . 2018. 高分四号卫星在干旱遥感监测中的应用 . 遥感学报, 22(3): 400-407) [DOI: 10.11834/jrs.20187067]

Pan T. 2015. Technical characteristics of the GaoFen-2 satellite. Aerospace China, (1): 3-9 (潘腾 . 2015. 高分二号卫星的技术特点 . 中国航天, (1): 3-9)

Roy D P, Wulder M A, Loveland T R, Woodcock C E, Allen R G, Anderson M C, Helder D, Irons J R, Johnson D M, Kennedy R, Scambos T A, Schaaf C B, Schott J R, Sheng Y, Vermote E F, Belward A S, Bindschadler R, Cohen W B, Gao F, Hipple J D, Hostert P, Huntington J, Justice C O, Kilic A, Kovalskyy V, Lee Z P, Lymburner L, Masek J G, McCorkel J, Shuai Y, Trezza R, Vogelmann J, Wynne R H and Zhu Z. 2014. Landsat-8: science and product vision for terrestrial global change research. Remote sensing of Environment, 145: 154-172 [DOI: 10.1016/j. rse. 2014. 02.001]

Shao Z F, Fu H Y, Li D R, Altan O and Cheng T. 2019. Remote sensing monitoring of multi-scale watersheds impermeability for urban hydrological evaluation. Remote Sensing of Environment, 232: 111338 [DOI: 10.1016/j.rse.2019.111338]

Wang L Z, Ma Y, Yan J N, Chang V and Zomaya A Y. 2018. pipsCloud: high performance cloud computing for remote sensing big data management and processing. Future Generation Computer Systems, 78: 353-368 [DOI: 10.1016/j.future.2016.06.009]

Wang Q, Wu C Q, Li Q and Li J S. 2010. Chinese HJ-1A/B satellites and data characteristics. Science China Earth Sciences, 53(1): 51-57 [DOI: 10.1007/s11430-010-4139-0]

Yan J N, Ma Y, Wang L Z, Choo K K R and Jie W. 2018. A cloud-based remote sensing data production system. Future Generation Computer Systems, 86: 1154-1166 [DOI: 10.1016/j. future. 2017. 02.044]

Yang J, Dong C H, Lu N M, Yang Z D, Shi J M, Zhang P, Liu Y J and Cai B. 2009. FY-3A: the new generation polar-orbiting meteorological satellite of china. Acta Meteorologica Sinica, 67(4): 501-509 (杨军, 董超华, 卢乃锰, 杨忠东, 施进明, 张鹏, 刘玉洁, 蔡斌 . 2009. 中国新一代极轨气象卫星——风云三号 . 气象学报, 67(4): 501-509) [DOI: 10.11676/qxxb2009.050]

Zhang Y J, Zheng M T, Xiong J X, Lu Y H and Xiong X D. 2014. On-Orbit geometric calibration of ZY-3 three-line array imagery with multistrip data sets. IEEE Transactions on Geoscience and Remote Sensing, 52(1): 224-234 [DOI: 10.1109/TGRS.2013.2237781]

Zhang Z, Tang P, Li H Y and Feng Z. 2016. Refined domain model for multisource data synergized quantitative remote sensing production system. Journal of Remote Sensing, 20(2): 184-196 (张正, 唐娉, 李宏益, 冯峥 . 2016. 多源数据协同定量遥感产品生产系统的领域模型 . 遥感学报, 20(2): 184-196) [DOI: 10.11834/jrs.20164293]

Zhong B, Yang A X, Liu Q H, Wu S L, Shan X J, Mu X H, Hu L F and Wu J J. 2021. Analysis ready data of the Chinese GaoFen satellite data. Remote Sensing, 13(9): 1709 [DOI: 10.3390/rs13091709]

Zhou J M, Zhang X, Liu Z P and Li Z. 2021. Extraction and analysis of mountain glacier movement from GF-1 satellite data. National Remote Sensing Bulletin, 25(2): 530-538 (周建民, 张鑫, 刘志平, 李震 . 2021. 高分一号山地冰川运动速度提取与分析 . 遥感学报, 25(2): 530-538) [DOI: 10.11834/jrs.20219080]

GF quantitative remote sensing production system: Core design

ZHANG Zheng, LI Hongyi, HU Changmiao, TANG Ping

Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

Abstract: A fast development in remote sensing science and technologies has been witnessed in recent years with the launch of various remote sensing satellites and the establishment of numerous industrial application systems. Satellite series, like GaoFen, has pushed the

richness of data to a new level. Moreover, quantitative product-driven application systems have become increasingly influential in many disciplines. By contrast, the bridge between data and application, namely, the production capability of quantitative products, seems weak, greatly limiting the usage and influence of GaoFen data. The improvement of production capability comes from multiple aspects, including product hierarchy, algorithm model, and production system. In this study, we propose, from the production system perspective, an integrated system design that responds to the four main challenges of the system: uniform data access, heterogeneous algorithm integration, layered workflow orchestration, and cloud infrastructure adaptation.

The system is based on algorithm containerization, where executable algorithms and all their dependencies are encapsulated. Thus, it can run uniformly and consistently on different infrastructures without worrying about the complexity caused by deployment. This scenario helps the system to manage diverse remote sensing algorithms uniformly. We employ the Kubernetes container orchestration platform to automate the execution, scaling, and management of containerized algorithms. A containerized cluster consists of multiple master nodes, many computing nodes, and multiple data centers. Multiple algorithm repositories are constructed to support the system and cope with the high computing and data throughput density of remote sensing algorithms. Each algorithm repository is further divided into several subrepositories to improve load balancing. User-defined role-based access control for algorithms is set up to protect the intellectual properties of algorithm owners. A recommended algorithm image architecture is introduced to standardize algorithm encapsulation. A set of nine properties are abstracted to describe uniformly any data entity parameter of an algorithm. This approach ensures that suitable input data can be found for user-uploaded algorithms to run in the system. For data visualization and quantitative computing scenarios, a multiscenario data organization strategy is proposed to avoid excessive data operations, such as projection transform or subdivision. The business logic of the system, from user order creation to product calculation, is detailed for clear implementation. The production sometimes involves workflow batches. We propose a stratified workflow aggregation strategy to optimize workflow execution.

The system has been used for large-scale production of various GF quantitative remote sensing products, including surface reflectance product, normalized difference vegetation index product, leaf area index product, and surface albedo product. These products fully cover China's area for eight successive years from 2013 to 2020, with quantities of more than five million and storages of nearly 300 TB. The proposed system completes the production task smoothly and efficiently.

During the routine support for many large-scale production tasks, each part of the system performed consistently with the system design proposed in this study, demonstrating that the study can help build a stable and efficient quantitative remote sensing production system on cloud-native infrastructures

Key words: production system, quantitative product, algorithm integration, container, cloud computing, system architecture, workflow

Supported by National Key Research and Development Program of China (No. 2018YFA0605500); National Natural Science Foundation of China (No. 41701399); China High-resolution Earth Observation System (No. 21-Y20B02-9003-19/22)