

时空遥感云计算平台 PIE-Engine Studio 的研究与应用

程伟, 钱晓明, 李世卫, 马海波, 刘东升, 刘富乾, 梁军龙, 胡举

航天宏图信息技术股份有限公司, 北京 100195

摘要: 随着遥感大数据时代的到来, 为快速处理和分析海量遥感数据, 国内外涌现了众多遥感云计算平台, 使得全球尺度、长时间序列遥感数据的快速分析和应用成为可能。本文在分析国内外遥感云计算平台现状的基础上, 针对大数据时代国内缺少功能完备的遥感云计算平台, 且国外遥感云计算平台对国产卫星数据支持不足等问题, 基于容器云技术, 构建了包含国产卫星数据且集数据、算力和技术于一体的时空遥感云计算平台 PIE (Pixel Information Expert) -Engine Studio, 实现了脚本驱动的遥感数据的按需获取以及海量数据的快速处理。采用 Landsat 8 数据, 以生长季植被指数 NDVI (Normalized Difference Vegetation Index) 的计算为例, 对比了本平台与 GEE (Google Earth Engine) 的数据处理能力。结果表明, 由于计算资源的限制, 本平台的计算和导出时间均比 GEE 稍长, 但计算结果的空间分布一致, 其中近 68% 的值均分布在 (0.48, 0.77), 且二者差值的 95.33% 集中在 (-0.13, 0.13), 结果较为可信。因此, 本文构建的基于共享、开放的中国自主遥感云计算平台 PIE-Engine Studio, 可为地球科学领域的研究提供数据和算力支持, 将有助于推进中国遥感云计算平台的发展进程, 推动国产卫星数据在云计算平台上的应用。

关键词: 遥感, 大数据, 遥感云计算平台, 分布式存储, 并行计算

引用格式: 程伟, 钱晓明, 李世卫, 马海波, 刘东升, 刘富乾, 梁军龙, 胡举. 2022. 时空遥感云计算平台 PIE-Engine Studio 的研究与应用. 遥感学报, 26(2): 335-347

Cheng W, Qian X M, Li S W, Ma H B, Liu D S, Liu F Q, Liang J L and Hu J. 2022. Research and application of PIE-Engine Studio for spatiotemporal remote sensing cloud computing platform. National Remote Sensing Bulletin, 26(2): 335-347 [DOI: 10.11834/jrs.20211248]

1 引言

随着对地观测技术的发展, 海量遥感数据不断涌现, 遥感大数据时代已经到来 (李德仁等, 2014; 李德仁, 2016; 付东杰等, 2021)。海量数据的存储、处理与共享对计算机性能提出了较高要求, 需要大量的存储与计算资源。云计算基于“按需分配”和“共享资源”的理念, 为海量遥感数据的存储、快速处理和分析提供了可能, 遥感专用的云计算平台也应运而生 (付东杰等, 2021)。遥感云计算平台的出现直接改变了传统遥感数据的处理和分析模式 (数据下载、存储和运算均在本地进行, 数据处理分析依赖本地运行的专业软件或代码), 降低了遥感数据的使用门槛,

极大地提高了运算效率, 使得全球尺度、高分辨率、长时间序列数据的快速分析和应用成为可能, 已成为研究土地利用、生态、环境和气候变化等地学领域前沿问题的重要工具 (董金玮等, 2020)。

目前国际上主流的遥感云计算平台以国外机构或公司研发的平台为主, 如美国的谷歌地球引擎 GEE (<https://earthengine.google.com/> [2021-04-29]) (Moore 和 Hansen, 2011)、美国航空航天局的 NEX (NASA Earth Exchange) (<https://www.nasa.gov/nex/> [2021-04-29]) (Nemani 等, 2010)、笛卡儿实验室的 Descartes Labs (<https://www.descarteslabs.com/> [2021-04-29])、亚马逊的 AWS (Amazon Web Services) (<https://aws.amazon.com/cn/earth/> [2021-04-

收稿日期: 2021-04-29; 预印本: 2021-12-06

第一作者简介: 程伟, 1987年生, 女, 博士, 研究方向为环境变化监测。E-mail: chengwei@piesat.cn

通信作者简介: 刘东升, 1984年生, 男, 高级工程师, 研究方向为遥感图像处理与应用。E-mail: liudongsheng@piesat.cn

本文算法代码共享: <https://engine.piesat.cn/engine-share/shareCode.html?id=07973edc3e8d4fc9af037d60987a5040>

29])、澳大利亚的数据立方体 Data Cube (<https://www.opendatacube.org/>[2021-04-29]) 以及德国的 CODE-DE (Copernicus Data and Exploitation Platform-DE) (<https://code-de.org/>[2021-04-29])。国内遥感云计算平台的建设也在快速推进中, 如中国科学院的地球大数据挖掘分析系统 EarthDataMiner (<http://earthdataminer.casearth.cn/>[2021-04-29]) (Liu等, 2020a), 多家科技公司如阿里巴巴集团控股有限公司、深圳市腾讯计算机系统有限公司、北京市商汤科技开发有限公司、航天宏图信息技术股份有限公司等也相继投入到遥感云计算平台的研发中。

在这些平台中, 以 GEE (Gorelick等, 2017) 的应用最为广泛, 在土地覆被/土地利用 (Bastin等, 2017; Ceccherini等, 2020; Finer等, 2018; Hansen等, 2013, 2020; Qin等, 2019)、植被 (Bastin等, 2019; Bogard等, 2019; Laskin等, 2019; Liu等, 2020b; Myers-Smith等, 2020)、动物 (Giezendanner等, 2020; Hendershot等, 2020; Joshi等, 2016)、气候变化 (Chudley等, 2019; Ryan等, 2019; Tuckett等, 2019; Venter等, 2020)、城市 (Liu等, 2020c)、农业 (Burke和Lobell, 2017)、灾害 (Walter等, 2019)、人类社会经济活动 (Watmough等, 2019; Weiss等, 2018; Wu等, 2020; Yeh等, 2020) 等地球科学领域中已得到典型应用。

GEE 历经近 10 a 发展, 技术较为成熟, 全球用户达数百万, 在此领域已处于垄断地位。然而美国政府于 2020-01 在出口管制条例上增加了“专门用于自动分析地理空间图像的软件” (Bureau of Industry and Security, Commerce, 2020), 提高了从美国向除加拿大之外的国家出口这些软件的限制, 而 GEE 正符合这个限制条件, 可能会被限制使用。目前, GEE 已提高了注册门槛, 国内新用户正常使用该平台越来越难, 并且使用过程及处理结果的下载受网络情况影响较大。同时由于 GEE 没有接入高分、风云等国产卫星数据, 限制了国产遥感数据在云计算平台上的应用。此外, 国内的遥感云计算平台目前主要以可视化和数据服务为主, 不能满足遥感行业对集数据、算力和技术于一体的遥感云计算平台的迫切需要。在这样的背景下, 航天宏图针对大数据时代遥感应用在国内面临的问题, 自主研发 PIE-Engine Studio 时空遥感云计算平台, 构建共享数据、算法和计算资源的开放环境, 实现遥感数据的按需获取以及海量数据的快

速处理, 以应对遥感大数据时代对信息高效、服务快速的迫切需求, 加速推进中国遥感云计算平台的发展进程。

2 PIE-Engine Studio 系统架构

PIE-Engine Studio 是一套基于容器云技术构建的面向地球科学领域的时空遥感云计算平台, 其内部包含自动管理的弹性大数据环境, 集成了多源遥感数据处理、分布式资源调度、实时计算、批量计算和深度学习框架等技术。作为构建在云计算之上的地理空间数据分析和计算平台, PIE-Engine Studio 通过结合海量遥感数据及计算资源, 可供用户在任意尺度上研究算法模型并采取交互式编程验证, 实现快速探索地表特征, 发现变化和趋势, 为地球科学领域的研究提供开放的数据与弹性算力支持。

PIE-Engine Studio 的系统架构如图 1 所示, 平台提供了 JavaScript 和 Python 两种版本的 API, 使用时在 Web 开发编辑器 (JavaScript) 或者 Jupyter Notebook (Python) 中编写脚本, 来处理分析各种影像、矢量数据。在 API 库的支持下, RESTful API 接口负责前端和后台的通信, 既可将前端脚本发送给后台, 又可将后台计算结果返回给前端显示, 也可将处理结果加载到第三方应用程序中进行展示或进一步分析。分布式数据计算服务负责对脚本进行编译, 对计算过程进行优化, 对计算计划进行执行并将计算结果通过 API 接口返回, 完成海量数据的分布式计算。平台提供的分布式数据计算服务分为两种: (1) 实时计算服务, 主要负责将计算结果实时显示到前端地图界面或者输出窗口; (2) 异步计算服务, 主要负责将数据以任务的形式导出。计算服务的执行依赖于平台分布式存储的多源遥感数据以及对这些遥感数据进行处理的算子库。分布式数据存储与访问服务负责多源数据的存储与访问, 包括以瓦片形式存储的影像数据、矢量要素数据、元数据以及用户个人数据。遥感数据处理分析算子库则包括了常用的遥感影像处理分析算法, 通过这些算子, 可以进行简单的数据处理以及复杂的图像运算。

通过平台主页 (<https://engine.piesat.cn/engine-studio/>[2021-09-01]) 可进入编程界面, 并查看数据资源、示例代码和文档教程等入门资源。平台也配备了免费的资源配额以上传个人数据或保存计算结果 (250 GB 存储空间, 6000 个文件), 亦可

根据需要对计算结果进行下载(60 GB/月免费下载流量)。另外,为保证计算资源的均衡分配,每次计算最多可免费调用2000景影像,导出结果大小限制为 10^{10} 个像素。



图1 PIE-Engine Studio 遥感云计算平台系统架构

Fig.1 The system architecture diagram of PIE-Engine Studio

3 分布式数据存储与访问

3.1 数据资源

截止到2021-09,平台已上传总数据量接近

3.4 PB,包括国内外100多种公共数据集(表1),近900万景影像,并以7 TB/日数据量持续更新。这些数据集涵盖了地球科学领域研究中常用的卫星影像以及气候、高程、土壤、土地覆盖、人口等数据产品,还包括了全球国家、中国省/市/县行政区划等基础矢量数据。所有数据经过投影转换、格式转换等预处理,可随时导入到代码编辑器中进行处理与分析。同时,用户可在资源目录中上传个人数据,并根据需要与其他用户分享。平台数据集访问地址为 <https://engine.piesat.cn/engine-studio/dataset> [2021-09-01]。

3.2 分布式数据存储

分布式数据存储首先要解决分布在国内外的海量多源遥感数据入库问题。对于国外免费遥感卫星数据,通过建立在AWS上的中转服务集群将数据实时同步到国内云数据中心;国内的遥感数据则直接传输到云数据中心。数据入库流程如图2所示,首先对分布在云端的海量遥感数据根据数据集名称、时间以及入库区域等信息进行筛选,生成需要入库的任务清单;之后将任务清单分配到不同的节点上并行执行,对任务清单中的每一条数据标识其状态信息,根据此标识采用分布式锁的抢占式任务调度方式保证入库数据的唯一性;最后将数据存储到云数据中心。

表1 PIE-Engine Studio 常用公共数据集

Table 1 Frequently used datasets in the PIE-Engine Studio data catalog

数据集	分辨率	时间粒度	时间覆盖	空间覆盖	
Landsat	Landsat 5/7/8 surface reflectance (C2)	30 m	16 d	1984年—现在	中国
	Landsat 8 surface reflectance (C1)	30 m	16 d	2013年—现在	中国
	Landsat 5 top of atmosphere (C2)	30 m	16 d	1984年—2012年	中国
	Landsat 7/8 top of atmosphere (C1)	30 m	16 d	1999年—现在	中国
Sentinel	Sentinel-1 A/Bground range detected	10 m	6 d	2020年—现在	中国
	Sentinel-2 MSI	10/20 m	10 d	2019年—现在	中国
	Sentinel-5P TROPOMI	0.01 弧度	5 min	2020年—现在	中国
MODIS	MOD09A1/MOD09Q1	500 m/250 m	8 d	2000年—现在	中国
	MOD09GA/MYD09GA	1 km, 500 m	每天	2000年—现在	中国
	MOD11A1/MOD11A2	1000 m	每天/8 d	2000年—现在	全球
	MCD12Q1	500 m	每年	2001年—2019年	全球
	MOD13A1/MOD13A2/MOD13Q1	250 m/500 m/1000 m	16 d	2000年—现在	中国
	MOD14A2	1 km	8 d	2000年—现在	全球
	MOD15A2H	500 m	8 d	2000年—2020年	中国
MOD16A2	500 m	8 d	2000年—现在	中国	

续表

	数据集	分辨率	时间粒度	时间覆盖	空间覆盖
高程	SRTM DEM	30 m	单次	2000年	中国
	ASTER GDEM	30 m	单次	2000年	中国
	ALOS DEM	12.5 m	单次	2011年	中国
夜光	DMSP_OLS	30 arcsec	每年	1992年—2013年	全球
	DMSP_OLS_RAD	30 arcsec	每年	1996年—2011年	全球
	VIIRS Nighttime lights	750 m	每月	2012年—现在	全球
气象	FY-4A	0.5—1 km、2—4 km	15 min	最近90 d	中国
	气象短临预报产品	0.05°	6 min	最近15 d	中国
	气象实况监测产品	0.05°	6 min	最近15 d	中国
	气象数值预报产品	0.05°	1 h	最近15 d	中国
	Himawari-8	2 km	10 min	最近30 d	中国
高分	GF-1_WFV_TOA	16 m	—	2019年	中国
	GF-1_WFV_SR	16 m	—	2019年	中国
	GF-1A/B/C/D多光谱地表反射率产品	8 m	—	2021年	中国
	GF-1A/B/C/D正射融合产品	2 m	部分按季度	2020年	中国部分区域
	GF-2正射融合产品	0.8 m	部分按季度	2020年	中国部分区域
	GF-6多光谱地表反射率产品	8 m	—	2021年	中国部分区域
资源	ZY-3 02多光谱地表反射率产品	8 m	—	2021年	中国部分区域
人口密度	WorldPop人口数据集	100 m/1 km	每年	2000年—2020年	中国
	LandScan全球人口分布数据集	1 km	每年	2000年—2017年	全球
国家青藏高原 科学数据中心	全球高分辨率地表太阳辐射产品	10 km	3 h	1983年—2017年	全国
	中国1 km分辨率月最低、高气温/平均气温、降水量产品	1 km	每月	1901年—2017年	中国
	长序列高时空分辨率月尺度温度和降水产品	0.025°	每月	1951年—2011年	中国
	全球长时间序列高分辨率生态系统总初级生产力产品	0.05°	每月	1982年—2018年	全球
	中国陆地实际蒸散发产品	0.1°	每月	1982年—2017年	全国
	全球PML_V2陆地蒸散发与总初级生产力产品	0.05°	8 d	2002年—2019年	全球
	中国土壤有机质产品	30 arcsec	单次	1980s	中国
	世界土壤数据库(HWSD)土壤产品(v1.2)	0.9°×1.25°	单次	2000年	全球
	中国土壤特征产品	1 km	单次	2010年	中国
	中国雪深长时间序列产品	25 km	每天	1979年—2019年	中国
土地覆被	全球地表覆盖产品 GlobeLand30	30 m	单次	2000年、2010年、 2020年	全球
	全球30 m精细地表覆盖产品	30 m	单次	2020年	全球
	全球10 m地表覆盖产品(FROM-GLC10 2017 v1)	10 m	单次	2017年	全球
	全球30 m地表覆盖产品(FROM-GLC30 2015 v2)	30 m	单次	2015年	全球
	全球30 m不透水面产品(MSMT_IS30)	30 m	单次	2015年	全球
	中国土地利用遥感监测产品	1 km	单次	1980年—2020年	中国
	中国建成区长时间序列产品	30 m	每年	1978年、1985年— 2017年	中国
WorldClim	全球2.5'分辨率最低/高气温、累积降水量产品	2.5 arcmin	每月	1961年—2018年	全球
指数产品	全球250 m LAI产品	250 m	8 d	2015年	全球
	全球1 km分辨率TVDI产品	1 km	每月	2000年—2020年	全球
全球历史 粮食产量	全球历史小麦/水稻/玉米/大豆产量产品	0.5°	每年	1981年—2016年	全球
矢量数据	全球国家行政区划边界		单次	2013年	全球
	中国省/市/县级行政区划		单次	2015年	中国

注：“—”表示仅空间上全覆盖,此表统计时间截止到2021-09,平台公共数据集持续更新中。

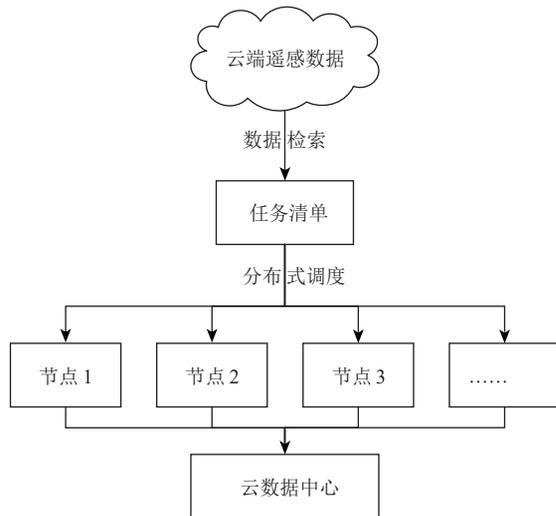


图2 分布式数据接入流程

Fig.2 The access process of distributed data

为支持海量遥感影像数据快速显示及高效计算,对影像数据建立影像金字塔。通过对影像重采样,建立一系列不同分辨率的影像图层,对每个图层分别进行切片处理得到影像瓦片,每个瓦片是大小为 256×256 像素的无损压缩数据块,并按照全球统一编码规则进行编码。对编码后的瓦片采用瓦片映射技术,构建瓦片之间的关联关系,然后采用公有云的对象存储技术,实现影像金字塔瓦片数据的分布式存储。

3.3 分布式数据访问

为影像建立了金字塔之后,基于云原生分布

式数据库构建影像数据集索引,便于通过脚本查询和过滤所需的影像。平台依据前端脚本代码传入的影像ID、当前交互区域的可见范围、当前地图的缩放级别等参数筛选出对应的影像瓦片,然后对这些瓦片进行处理。采用这种分块瓦片计算的处理方式,执行过程耗时短,效率高。由于获取的是多个单张瓦片,所以也便于将后台计算服务做分布式部署,通过分布式操作使得对海量数据的实时计算更为高效。

此外,为使用户轻松搜索并获取自己想要的数 据,平台把相同传感器拍摄的同一类型的数据组合在一起,成为数据集(collection)的形式(如:Landsat 8 TOA 数据集)。该种方式提供了快速过滤和排序功能,用户可以很容易地搜索数百万张单独的影像,以选择特定空间、时间或其他条件的数据。例如,可以轻松地从Landsat 8 TOA 数据集中查询华北区域2020年云量小于10%的数据并显示在地图界面上。

4 遥感数据处理算子

PIE-Engine Studio 提供了针对数值、数组/矩阵、影像、矢量、集合、列表、字典等多种对象类型的操作,包含机器学习分类算法和部分卫星的专用算法如气溶胶光学厚度反演、PM 2.5 反演、海表温度提取、舰船检测等,共计300多个算子,支持JavaScript和Python两种语言的访问方式,具体算子如表2所示。

表2 PIE-Engine Studio 支持的算子

Table 2 PIE-Engine Studio function summary

算法类型	算法说明	具体算子
数值运算	四则运算	Add,subtract,multiply,divide等
	三角函数	Sin,cos,tan,asin,acos,atan等
	标准运算	Abs,pow,sqrt,floor,min,max,round等
	逻辑运算	Eq,neq,gt,gte,lt,lte,and,or,not
	按位运算	Bitwise and/or/not/xor等
	类型转换	Int,double,float,byte等
数组/矩阵运算	元素运算	同数值运算
	数组操作	Get,length,cat,slice等
	矩阵运算	Determinant,decomposition,solve,transpose,inverse,diagonal,pseudoinverse等
	统计分析	Accum,reduce
影像操作	基本操作	同数值运算
	影像裁剪	Clip
	波段操作	Add,select,rename,expression,expression function等
	统计分析	Reduce,reduce region等
	随机采样	Sample,sample regions等

续表

算法类型	算法说明	具体算子
	地形分析	Slope、aspect、hillshade等
	元数据属性	Get、set等
	其他操作	Cat、constant、geometry、paint、pixel area、reproject、update mask、where、convolve等
影像集合	基本操作	Size、first、map、mosaic、sort等
	影像过滤	Filter、filter bounds、filter date、select、limit等
	统计分析	Max、mean、median、min、reduce、reduce columns、sum等
	元数据属性	Get、set等
几何操作	几何类型	Point、line string、polygon、rectangle等
	量测操作	Length、area等
	空间分析	Intersects、contains、containedin、disjoint、within distance、intersection、union、difference等
	其他操作	Buffer、bounds、centroid、coordinates、simplify等
矢量操作	量测操作	Length、area等
	空间分析	同几何操作
	元数据属性	Get、set等
	其他操作	Buffer、simplify、geometry等
矢量集合	基本操作	First、map、merge等
	过滤操作	Filter、filterbounds、limit等
	元数据属性	Get、set等
	其他操作	Geometries、error matrix、classify、random column等
聚合	简单聚合	Count、first、last等
	统计分析	Mean、median、mode、percentile、histogram等
	相关分析	Pearson correlation
	回归分析	Linear fit、linear regression
核运算	卷积核	Horizontal、laplacian、lowpass、sobel、vertical等
机器学习	监督分类	KNN、Random Forest、SVM、Normal Bayes等
	非监督分类	K-Means、EM
专用算法	葵花8号	AOT、cloud mask、rgb、PM 10、PM 2.5、sea surface temperature
	风云4A	AOT、cloud mask、rgb、PM 10、PM2.5、snow percent
	Landsat 8、Sentinel 2	Cloud mask
	Sentinel-1、高分三号	Coastline extract、ship detection
过滤操作	时间过滤	Date、day of year、calendar range
	空间过滤	Bounds、contains、within distance等
	属性过滤	In list、string contains、string starts/ends with
	逻辑过滤	Eq、neq、gt、gte、lt、lte、and、or、not、not null
精度评价	混淆矩阵	Accuracy、kappa
其他数据类型	字符串、列表、字典、地图、日期、投影等	

注：此表统计时间截止到2021-09，平台算子持续更新中。

上述算子根据复杂度的不同可分为简单算子和复杂算子，其中简单算子是指仅通过单块影像计算的算法，如波段运算或拉伸显示等。复杂算子是指需要统计整体影像的算法，如统计极值、直方图或计算面积等。

一个算子由类型、名称和参数列表3部分组成，其中类型用来标识是简单算子（Simple）还是复杂算子（Complex），名称即该算子的函数名，参数列表指该算子的参数信息。比如执行影像加法运算的算子：*addImage*（imageA，imageB），其

类型为Simple，名称为*addImage*，包含两个参数，分别为imageA和imageB。

使用时既可单独调用某个算子实现相应的功能，如用算子*Image*加载指定的影像：*var image = pie.Image ("LC08/01/T1/LC08_121031_20170101")*，又可通过对这些算子进行组合调用，来实现复杂的数据查询、计算、分析和显示等功能。如算子*Image*和算子*select*结合可实现对特定影像的波段筛选功能：*var image = pie.Image ("LC08/01/T1/LC08_121031_20170101").select ("B1")*；

平台所有算子均采用C++语言实现, 具有较高的计算效率。随着遥感数据的扩展和用户需求的不断丰富, 会对新的算子提出需求, 为此平台支持多语言的算子扩展机制, 包括 JavaScript、Python 和 C++。

5 分布式数据计算

PIE-Engine Studio 运行的概略流程是: 首先在前端编写动态脚本 (JavaScript/Python), 描述遥感影像的计算过程; 点击运行按钮, 这些 JavaScript/Python 代码自动构建初步的链式结构调用语法树 (即树结构的调用依赖关系); 然后通过后台的计算过程优化以过滤无效计算内容, 生成并缓存所需数据结果的执行计划; 接着通过调度中心将计算任务分配给多个节点上的计算服务; 最终通过特定的前端请求或者算子 (*print*、*addLayer*、*Export*) 触发实际的计算, 以分布式方式完成计算流程, 得到可视化图层或者结果文件返回给前端界面。这一流程如图3所示。

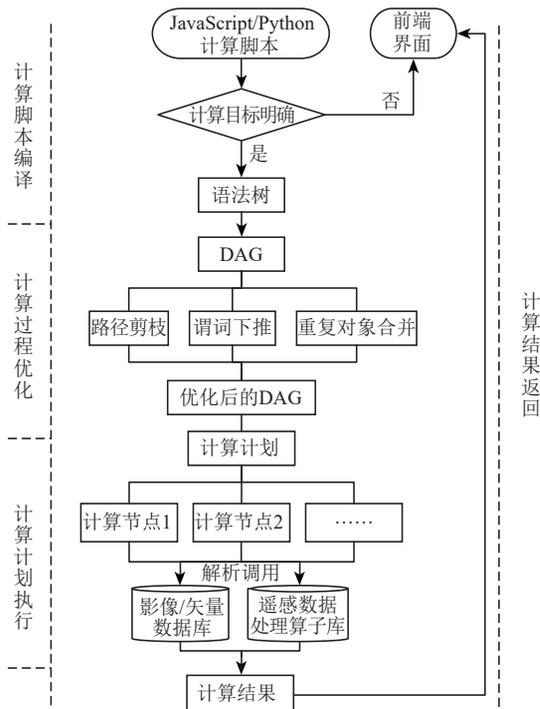


图3 PIE-Engine Studio 遥感云计算平台计算流程

Fig.3 The calculation flow of PIE-Engine Studio

5.1 计算脚本编译

PIE-Engine Studio 的执行是脚本驱动的, 脚本是在 Web 编辑器中编写或者通过 RESTful API 接口提交的文本内容, 因此需要转换为与后台算子

库匹配的 C++ 计算指令才可在众多服务器上并行执行。为达到这一目的, 平台设计了一套链式结构调用语法树以实现脚本语言的编译, 语法树中的节点是数据对象或者预定义的算子, 后期的计算过程优化和计算计划生成均依赖这个树结构。

PIE-Engine Studio 接收到动态脚本后, 首先识别计算的最终目标对象, 对于没有计算目标的脚本 (如仅操作前端 UI 部件) 则直接终止整个计算过程; 对于计算目标明确的脚本, 则执行计算脚本编译的任务, 生成链式结构调用语法树, 并通过触发计算的算子确定输出结果的形式。

(1) 单个算子的编译。单个算子的编译是通过前端脚本代码对单个算子进行构造, 获取算子的类型、名称以及参数列表, 实现该算子的语法树描述信息。

(2) 组合算子的编译。对于链式调用的组合算子, 首先前端通过算子类型、名称和参数列表对每个算子进行构造, 得到相应的语法树描述信息; 然后根据算子链式调用顺序把前一个算子的描述信息作为后一个算子的输入参数, 依次类推构建组合算子的语法树描述信息。

5.2 计算过程优化

为实现海量遥感数据批量处理, PIE-Engine Studio 计算逻辑十分复杂, 为提高运算效率, 对计算过程进行优化尤为重要。首先平台根据语法树的计算目标, 构建计算执行路径的有向无环图 DAG (Directed Acyclic Graph), 关于 DAG 的构建过程可参考 Gorelick 等 (2017) 的描述。然后采用多种技术对 DAG 进行优化, 主要包括: 通过计算路径剪枝, 移除不必要的计算过程; 通过谓词下推, 把对影像的过滤条件尽可能下推至靠近数据源的位置, 从源头减少计算所需的数据量; 通过重复对象合并, 将不同计算分支上的相同对象合并, 避免重复计算。在代码层次实现这些优化, 可有效提高平台执行速度和资源利用效率。最后根据优化后的 DAG 生成计算计划并缓存等待调用。

5.3 计算计划执行

(1) 计算任务解析。计算计划生成后需要转换为可执行的 C++ 调用函数才可与后台算子库匹配, 从而根据算子名称调用算子库中对应算子执

行相应的计算。计算服务通过解析语法树中每个变量包含的表达式, 获取表达式对应的具体内容, 然后通过递归方式逐步获取变量最终的C++调用函数。

解析单个算子的过程是通过算子的语法树描述信息, 解析出算子类型、名称和参数列表, 从而构建该算子的调用函数。解析链式调用的组合算子则先识别各个算子的嵌套关系。通过解析外部算子的参数, 判断该算子的参数是值还是其他算子的描述信息。如果是值则对该参数直接赋值, 如果是其他算子的描述信息则按照单个算子的解析过程构建该算子的C++调用函数。

PIE-Engine Studio采用的是按需计算, 也就是明确调用了上下文时程序才开始进行计算。例如只有在将计算结果进行输出 (*print*、*Export*) 或者添加到交互地图上 (*addLayer*) 时, 计算服务才会依据前端传入的影像ID、当前交互区域的可见范围、当前地图的缩放级别等, 查找所有对应的瓦片, 根据计算计划动态计算当前可见区域对应的结果。或者如果计算结果是另外一个计算过程的输入, 计算服务会对计算结果自动重新采样, 使得两者可以正常运算, 整个过程完全不需要用户控制分辨率。

在处理缓存问题上, 平台为节省计算资源, 对已经计算过的对象限时缓存 (5 min), 如果需要计算的对象存在缓存则直接提取计算结果替换对象的整个计算过程。这种方式既避免了重复计算降低效率, 又避免了长期无人访问的无效缓存占用存储空间。

(2) 计算任务调度。平台针对不同级别的任务提供了不同的调度策略, 主要有轮询调度 (Round Robin)、资源利用均衡、请求数均衡3种策略。其中轮询调度算法的主要思想是把用户每次提交的任务请求, 依次分配到内部的所有服务器中, 不关心每台服务器的当前连接数和响应速度。其他两种方式的目标分别是在各个计算服务器的计算资源和请求数之间取得平衡。当计算节点接收到调度的计算请求后, 根据请求中的脚本唯一标识到缓存中检索对应的执行计划并完成实际的计算。

(3) 并行计算策略。实时计算任务执行时, 其总执行计划与算子进行绑定, 实现多线程、并行IO的计算方式。针对复杂度不同的算子, 采取

的并行计算策略也不一样。对于简单算子, 仅通过单块影像即可计算, 采用直接计算的方式; 对于复杂算子, 即需要统计整体影像的算法, 采用MapReduce计算的方式; 对于更复杂的运算, 比如滤波运算, 需要先对单块影像的边界进行扩充, 再采用MapReduce的计算方式。

5.4 计算结果返回

对于实时计算服务, 直接将后台计算结果返回前端地图界面显示, 同时将需要输出的结果显示在输出窗口中。

对于异步计算服务, 则生成相关导出任务, 然后在后台异步执行直到任务导出结束, 相关结果会根据用户的设置导出到云盘或资源目录中。

6 应用案例

PIE-Engine Studio和GEE都属于遥感云计算平台, 两者在设计理念、开发方式、提供的服务等方面具有高度相似性, 作为中国自主研发的遥感云计算平台, 为在算力上与GEE作对比, 采用Landsat 8 TOA数据集, 以整体计算中国区域生长季植被指数均值为例来展示PIE-Engine Studio (以下简称本平台) 的数据处理能力。2020-05—2020-08, 全国共有3366景Landsat 8 TOA数据, 数据总量约3 TB。计算流程为: (1) 加载Landsat 8 TOA数据集; (2) 做去云处理; (3) 计算归一化植被指数NDVI; (4) 影像镶嵌裁剪得到中国区域生长季植被指数均值。对比结果归纳如表3所示。

6.1 计算时间的对比

在计算时间上, GEE需要0.5 min左右, 而本平台则需1—2 min, 主要原因是GEE背靠谷歌公司, 具有强大的后台计算服务做支撑; 而本平台囿于现阶段在公有云上的投入, 可供调用的计算服务有限, 因此在计算资源上与GEE存在差距, 导致目前本平台的计算时间比GEE稍长。

6.2 计算结果的对比

为进一步比较本平台与GEE的计算结果, 以北京市为研究区, 将两个平台2020年生长季NDVI的计算结果均以30 m分辨率导出, GEE的导出任务执行时间为6 min, 而本平台则需9 min。计算结果的空间分布分别如图4(a)、图4(b)所示, 二者差值如图4(c)所示。

表3 PIE-Engine Studio 和 GEE 计算结果对比
Table 3 Comparison of results calculated by Pie-Engine Studio and GEE

对比项	本平台	GEE	差别原因
2020年全国生长季 NDVI计算时间	1—2 min	0.5 min 左右	本平台计算服务资源有限,导致计算时间稍长
2020年北京市生长季 NDVI计算结果	大部分区域二者差值较小,仅在建筑密集区差值波动较大 直方图分布一致	由于重采样的原因,两平台导出数据在相同位置处存在错位,这种错位在建筑区比较明显,原因是建筑区像素间值差别较大,错位影响也就大,而边远山区下垫面较均一,错位影响较小 两平台计算结果中近68%的值均分布在(0.48,0.77),二者差值的95.33%集中在(-0.13,0.13)	
2020年北京市生长季 NDVI导出时间	9 min	6 min	本平台用于导出的计算服务资源有限,导致导出时间较长 两平台导出文件数据类型不同,本平台导出文件较大,导致导出时间增长

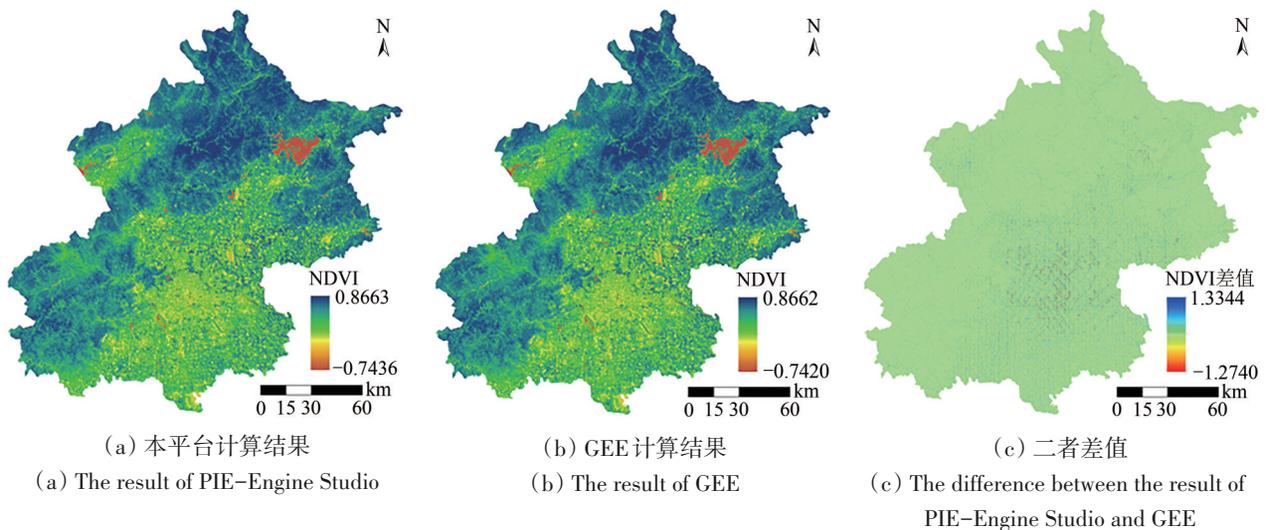


图4 2020年北京市生长季NDVI均值空间分布

Fig.4 The spatial distribution of mean NDVI in Beijing in 2020

从空间分布上看,本平台计算结果与GEE在空间分布上趋势一致,最大最小值也较为接近;在大部分区域,二者差值较小,仅在建筑密集区差值波动较大。造成这种现象的原因主要有两个:

(1) 两个平台存储Landsat 8 TOA时采用的数据类型不同。GEE采用的是Float32数据类型,而本平台为减少存储量及方便数据压缩,采用的是Int16类型(比例因子为0.0001)。数据类型的不同导致同一像素的数据精度不同,由波段运算得到的结果存在差别,但由图4可以看出,这种细微差别对计算结果的空间分布趋势未造成较大影响。

(2) 两个平台导出的数据在相同位置处存在像素错位。本平台在内部统一采用Web Mercator

(EPSG: 3857) 投影,在导出时,根据前端设置的分辨率,对导出数据最接近该分辨率的瓦片级别进行重采样,然后进行导出。在GEE中可对导出的数据进行坐标系设置,默认为WGS1984 (EPSG: 4326) 地理坐标系,当将导出分辨率设置为30 m时,在此坐标系下实际导出的数据分辨率比30 m要大($0.00026949459^\circ \approx 33.78523056294 \text{ m}$);当设置为EPSG: 3857时,实际导出的数据分辨率为30 m。但由于重采样的原因,两个平台导出的数据在相同位置处存在错位,像素值也不相同。这种错位在建筑区比较明显(图4(c)),原因是建筑区像素间值差别较大,错位带来的影响也就大;而在边远山区,下垫面比较均一,像素位置不重合带来

的影响较小。

分别统计两个平台导出结果及二者差值的直方图,结果如图5所示。可以看出两个平台导出结果的直方图分布趋势比较一致,分别有67.87% (本

平台)和67.72% (GEE) 的值分布在(0.48, 0.77)。从两个平台计算结果差值的直方图(图5(c))上可以看出,95.33%的值都集中在(-0.13, 0.13),因此从整体上看,本平台的计算结果较为可信。

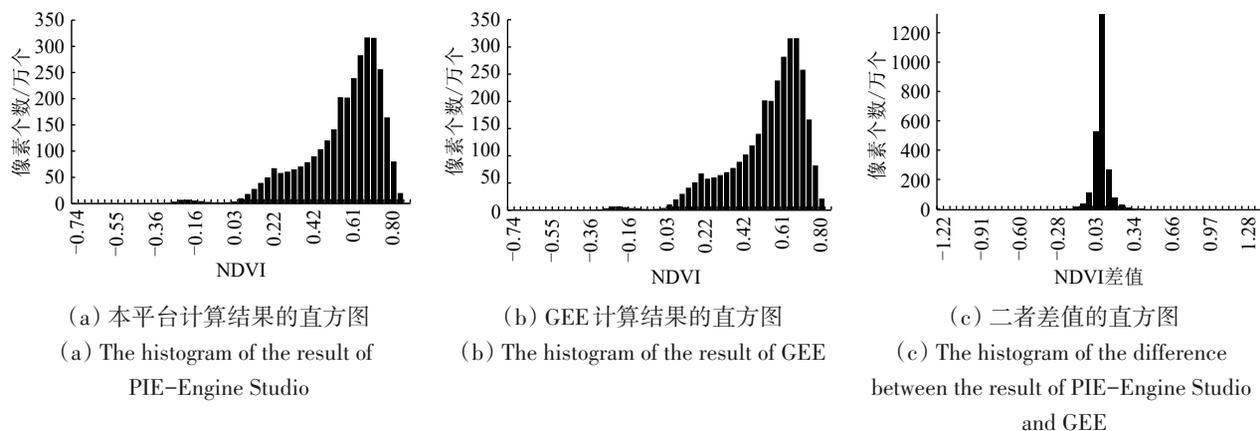


图5 2020年北京市生长季NDVI均值空间分布的直方图
Fig.5 The histogram of spatial distribution of mean NDVI in Beijing in 2020

此外,在导出时间方面,由于两个平台导出文件的数据类型并不相同,本平台默认导出的类型是Float64类型,GEE是Float32类型。理论上Float64类型占用的空间是Float32的2倍,但由于本平台未对导出数据进行压缩,实际数据大小约为GEE的4倍。因此除服务器资源的限制外,本平台导出的文件较大也会导致导出时间延长。

7 结论

在遥感大数据时代背景下,基于共享、开放的中国自主遥感云计算平台PIE-Engine Studio的研究与开发,将有助于推进中国遥感云计算平台的进展,提升国际影响力。与国外相比,目前中国的遥感云计算平台尚处于起步阶段,在数据、计算资源、算子、应用领域及智能化等方面尚存在不足,未来会进一步加大公有云的投入,扩充计算资源;继续研发新的算子,完善平台功能;持续入库卫星数据,丰富数据资源;加强院校合作,拓展应用领域;融合深度学习技术,提高智能化水平。但将国产卫星数据纳入遥感云计算平台,会让更多的用户参与和使用中国数据,推动国产卫星数据在云计算平台上的应用,在国内地球科学领域研究中将发挥重要作用,使得中国遥感云计算平台具有独特优势。随着云计算技术的发展以及人类进一步认识地球的需要,遥感云计算平

台将得到更多的应用和发展,为更深入理解地学规律,实现人类社会可持续发展提供强大的科学支撑。

参考文献(References)

- Bastin J F, Berrahmouni N, Grainger A, Maniatis D, Mollicone D, Moore R, Patriarca C, Picard N, Sparrow B, Abraham E M, Aloui K, Atesoglu A, Attore F, Bassüllü Ç, Bey A, Garzuglia M, García-Montero L G, Groot N, Guerin G, Laestadius L, Lowe A J, Mamane B, Marchi G, Patterson P, Rezende M, Ricci S, Salcedo I, Diaz A S P, Stolle F, Surappaeva V and Castro R. 2017. The extent of forest in dryland biomes. *Science*, 356(6338): 635-638 [DOI: 10.1126/science.aam6527]
- Bastin J F, Finegold Y, Garcia C, Mollicone D, Rezende M, Routh D, Zohner C M and Crowther T W. 2019. The global tree restoration potential. *Science*, 365(6448): 76-79 [DOI: 10.1126/science.aax0848]
- Bogard M J, Kuhn C D, Johnston S E, Striegl R G, Holtgrieve G W, Dornblaser M M, Spencer R G M, Wickland K P and Butman D E. 2019. Negligible cycling of terrestrial carbon in many lakes of the arid circumpolar landscape. *Nature Geoscience*, 12(3): 180-185 [DOI: 10.1038/s41561-019-0299-5]
- Bureau of Industry and Security, Commerce. 2020. Addition of software specially designed to automate the analysis of geospatial imagery to the export control classification number 0Y521 series[EB/OL]. (2020-01-06)[2021-04-29]. <https://www.federalregister.gov/documents/2020/01/06/2019-27649/addition-of-software-specially-designed-to-automate-the-analysis-of-geospatial-imagery-to-the-export>

- Burke M and Lobell D B. 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proceedings of the National Academy of Sciences of the United States of America*, 114(9): 2189-2194 [DOI: 10.1073/pnas.1616919114]
- Ceccherini G, Duveiller G, Grassi G, Lemoine G, Avitabile V, Pilli R and Cescatti A. 2020. Abrupt increase in harvested forest area over Europe after 2015. *Nature*, 583(7814): 72-77 [DOI: 10.1038/s41586-020-2438-y]
- Chudley T R, Christoffersen P, Doyle S H, Bougamont M, Schoonman C M, Hubbard B and James M R. 2019. Supraglacial lake drainage at a fast-flowing Greenlandic outlet glacier. *Proceedings of the National Academy of Sciences of the United States of America*, 116(51): 25468-25477 [DOI: 10.1073/pnas.1913685116]
- Dong J W, Li S W, Zeng Y L, Yan K and Fu D J. 2020. Remote Sensing Cloud Computing and Scientific Analysis-Application and Practice. Beijing: Science Press: 10-18 (董金玮, 李世卫, 曾也鲁, 闫凯, 付东杰. 2020. 遥感云计算与科学分析——应用与实践. 北京: 科学出版社: 10-18)
- Finer M, Novoa S, Weisse M J, Petersen R, Mascaro J, Souto T, Stearns F and Martinez R G. 2018. Combating deforestation: from satellite to intervention. *Science*, 360(6395): 1303-1305 [DOI: 10.1126/science.aat1203]
- Fu D J, Xiao H, Su F Z, Zhou C H, Dong J W, Zeng Y L, Yan K, Li S W, Wu J, Wu W Z and Yan F Q. 2021. Remote sensing cloud computing platform development and Earth science application. *National Remote Sensing Bulletin*, 25(1): 220-230 (付东杰, 肖寒, 苏奋振, 周成虎, 董金玮, 曾也鲁, 闫凯, 李世卫, 吴进, 吴文周, 颜凤芹. 2021. 遥感云计算平台发展及地球科学应用. *遥感学报*, 25(1): 220-230) [DOI: 10.11834/jrs.20210447]
- Giezendanner J, Pasetto D, Perez-Saez J, Cerrato C, Viterbi R, Terzagio S, Palazzi E and Rinaldo A. 2020. Earth and field observations underpin metapopulation dynamics in complex landscapes: near-term study on carabids. *Proceedings of the National Academy of Sciences of the United States of America*, 117(23): 12877-12884 [DOI: 10.1073/pnas.1919580117]
- Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D and Moore R. 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202: 18-27 [DOI: 10.1016/j.rse.2017.06.031]
- Hansen A J, Burns P, Ervin J, Goetz S J, Hansen M, Venter O, Watson J E M, Jantz P A, Virnig A L S, Barnett K, Pillay R, Atkinson S, Supples C, Rodríguez-Buritica S and Armenteras D. 2020. A policy-driven framework for conserving the best of Earth's remaining moist tropical forests. *Nature Ecology and Evolution*, 4(10): 1377-1384 [DOI: 10.1038/s41559-020-1274-7]
- Hansen M C, Potapov P V, Moore R, Hancher M, Turubanova S A, Tyukavina A, Thau D, Stehman S V, Goetz S J, Loveland T R, Kommareddy A, Egorov A, Chini L, Justice C O and Townshend J R G. 2013. High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160): 850-853 [DOI: 10.1126/science.1244693]
- Hendershot J N, Smith J R, Anderson C B, Letten A D, Frishkoff L O, Zook J R, Fukami T and Daily G C. 2020. Intensive farming drives long-term shifts in avian community composition. *Nature*, 579(7799): 393-396 [DOI: 10.1038/s41586-020-2090-6]
- Joshi A R, Dinerstein E, Wikramanayake E, Anderson M L, Olson D, Jones B S, Seidensticker J, Lumpkin S, Hansen M C, Sizer N C, Davis C L, Palminteri S and Hahn N R. 2016. Tracking changes and preventing loss in critical tiger habitat. *Science Advances*, 2(4): e1501675 [DOI: 10.1126/sciadv.1501675]
- Laskin D N, McDermid G J, Nielsen S E, Marshall S J, Roberts D R and Montaghi A. 2019. Advances in phenology are conserved across scale in present and future climates. *Nature Climate Change*, 9(5): 419-425 [DOI: 10.1038/s41558-019-0454-4]
- Li D R. 2016. Towards geo-spatial information science in big data era. *Acta Geodaetica et Cartographica Sinica*, 45(4): 379-384 (李德仁. 2016. 展望大数据时代的地球空间信息学. *测绘学报*, 45(4): 379-384) [DOI: 10.11947/j.AGCS.2016.20160057]
- Li D R, Zhang P L and Xia G S. 2014. Automatic analysis and mining of remote sensing big data. *Acta Geodaetica et Cartographica Sinica*, 43(12): 1211-1216 (李德仁, 张良培, 夏桂松. 2014. 遥感大数据自动分析与数据挖掘. *测绘学报*, 43(12): 1211-1216) [DOI: 10.13485/j.cnki.11-2089.2014.0187]
- Liu J, Wang W and Zhong H. 2020a. EarthDataMiner: a cloud-based big earth data intelligence analysis platform. *IOP Conference Series: Earth and Environmental Science*, 509: 012032 [DOI: 10.1088/1755-1315/509/1/012032]
- Liu Y L, Kumar M, Katul G G, Feng X and Konings A G. 2020b. Plant hydraulics accentuates the effect of atmospheric moisture stress on transpiration. *Nature Climate Change*, 10(7): 691-695 [DOI: 10.1038/s41558-020-0781-5]
- Liu X P, Huang Y H, Xu X C, Li X C, Li X, Ciais P, Lin P R, Gong K, Ziegler A D, Chen A P, Gong P, Chen J, Hu G H, Chen Y M, Wang S J, Wu Q S, Huang K N, Estes L and Zeng Z Z. 2020c. High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015. *Nature Sustainability*, 3(7): 564-570 [DOI: 10.1038/s41893-020-0521-x]
- Moore R T and Hansen M C. 2011. Google Earth Engine: a new cloud-computing platform for global-scale earth observation data and analysis//*Proceedings of the American Geophysical Union, Fall Meeting 2011*. [s.l.]: AGU
- Myers-Smith I H, Kerby J T, Phoenix G K, Bjerke J W, Epstein H E, Assmann J J, John C, Andreu-Hayles L, Angers-Blondin S, Beck P S A, Berner L T, Bhatt U S, Bjorkman A D, Blok D, Bryn A, Christiansen C T, Cornelissen J H C, Cunliffe A M, Elmendorf S C, Forbes B C, Goetz S J, Hollister R D, de Jong R, Loranty M M, Macias-Fauria M, Maseyk K, Normand S, Olofsson J, Parker T C, Parmentier F J W, Post E, Schaeppman-Strub G, Stordal F,

- Sullivan P F, Thomas H J D, Tømmervik H, Treharne R, Tweedie C E, Walker D A, Wilmking M and Wipf S. 2020. Complexity revealed in the greening of the Arctic. *Nature Climate Change*, 10(2): 106-117 [DOI: 10.1038/s41558-019-0688-1]
- Nemani R R, Votava P, Michaelis A, Melton F S, Hashimoto H, Milesi C, Wang W and Ganguly S. 2010. NASA earth exchange: a collaborative earth science platform//Proceedings of the American Geophysical Union, Fall Meeting 2010. [s.l.]: AGU
- Qin Y W, Xiao X M, Dong J W, Zhang Y, Wu X C, Shimabukuro Y, Arai E, Biradar C, Wang J, Zou Z H, Liu F, Shi Z, Doughty R and Moore III B. 2019. Improved estimates of forest cover and loss in the Brazilian Amazon in 2000-2017. *Nature Sustainability*, 2(8): 764-772 [DOI: 10.1038/s41893-019-0336-9]
- Ryan J C, Smith L C, van As D, Cooley S W, Cooper M G, Pitcher L H and Hubbard A. 2019. Greenland Ice Sheet surface melt amplified by snowline migration and bare ice exposure. *Science Advances*, 5(3): (eaav3738 [DOI: 10.1126/sciadv.aav3738])
- Tuckett P A, Ely J C, Sole A J, Livingstone S J, Davison B J, van Wessem J M and Howard J. 2019. Rapid accelerations of Antarctic Peninsula outlet glaciers driven by surface melt. *Nature Communications*, 10(1): 4311 [DOI: 10.1038/s41467-019-12039-2]
- Venter Z S, Aunan K, Chowdhury S and Lelieveld J. 2020. COVID-19 lockdowns cause global air pollution declines. *Proceedings of the National Academy of Sciences of the United States of America*, 117(32): 18984-18990 [DOI: 10.1073/pnas.2006853117]
- Walter T R, Haghighi M H, Schneider F M, Coppola D, Motagh M, Saul J, Babeyko A, Dahm T, Troll V R, Tilmann F, Heimann S, Valade S, Triyono R, Khomarudin R, Kartadinata N, Laiolo M, Massimetti F and Gaebler P. 2019. Complex hazard cascade culminating in the Anak Krakatau sector collapse. *Nature Communications*, 10(1): 4339 [DOI: 10.1038/s41467-019-12284-5]
- Watmough G R, Marcinko C L J, Sullivan C, Tschirhart K, Mutuo P K, Palm C A and Svenning J C. 2019. Socioecologically informed use of remote sensing data to predict rural household poverty. *Proceedings of the National Academy of Sciences of the United States of America*, 116(4): 1213-1218 [DOI: 10.1073/pnas.1812969116]
- Weiss D J, Nelson A, Gibson H S, Temperley W, Peedell S, Lieber A, Hancher M, Poyart E, Belchior S, Fullman N, Mappin B, Dalrymple U, Rozier J, Lucas T C D, Howes R E, Tusting L S, Kang S Y, Cameron E, Bisanzio D, Battle K E, Bhatt S and Gething P W. 2018. A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, 553(7688): 333-336 [DOI: 10.1038/nature25181]
- Wu X, Braun D, Schwartz J, Kioumourtzoglou M A and Dominici F. 2020. Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly. *Science Advances*, 6(29): (eaba5692 [DOI: 10.1126/sciadv.aba5692])
- Yeh C, Perez A, Driscoll A, Azzari G, Tang Z Y, Lobell D, Ermon S and Burke M. 2020. Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications*, 11(1): 2583 [DOI: 10.1038/s41467-020-16185-w]

Research and application of PIE-Engine Studio for spatiotemporal remote sensing cloud computing platform

CHENG Wei, QIAN Xiaoming, LI Shiwei, MA Haibo, LIU Dongsheng, LIU Fuqian, LIANG Junlong, HU Ju

Piesat Information Technology Co., Ltd., Beijing 100195, China

Abstract: With the arrival of remote sensing big data era, numerous remote sensing cloud computing platforms have emerged inland and overseas to rapidly process and analyze massive remote sensing data. The emergence of remote sensing cloud computing platform makes it possible to quickly analyze and apply remote sensing data on a global scale or for longterm sequences. However, currently, there is lacking of remote sensing cloud computing platform with complete functions in domestic, while foreign remote sensing cloud computing platform has insufficient support for domestic satellite data. Based on this situation, we have independently developed a spatiotemporal remote sensing cloud computing platform, PIE (Pixel Information Expert) -Engine Studio. By adopting container cloud technology, this platform integrating data, computing power and technology, can implements on-demand acquisition of remote sensing data and rapid processing of massive data just driven by the script. (1) This study first introduced the system architecture of PIE-Engine Studio, and then described the data storage and access mode. (2) PIE-Engine Studio provides operations for multiple objects such as number, matrix, image, vector, list, dictionary, etc., also machine learning algorithms and some special satellite algorithms. (3) Furthermore, this study illustrated the calculation flow of the platform in detail. Firstly, the user writes a script in the front-end to describe the calculation process of remote sensing data.

Click the “Run” button, these codes automatically build the preliminary chained structure call syntax tree. Then the syntax tree is optimized in the back-end through filter the invalid calculation content. The computing tasks are then distributed to the computing services on multiple nodes through the scheduling center. Finally, the resulting visual map layer or data file is returned to the front-end interface triggered by specific front-end requests or operators (print, addLayer, export).(4) At last, an application case is presented, we adopted Landsat 8 data and taking the calculation of Normalized Difference Vegetation Index (NDVI) in the growing season as an example, the calculation results and running time of this platform are compared with Google Earth Engine (GEE). The results show that, due to the limitation of computing resources, the running and export time of this platform are slightly longer than that of GEE, but the spatial distribution of calculation results is consistent, among which about 68% values are distributed between (0.48, 0.77), and 95.33% of the difference between the two results is concentrated between (-0.13, 0.13). It shows that the results are reliable. Therefore, the remote sensing cloud computing platform constructed by this paper, can provide data resources and computing power for research in the field of earth science, and will help promote the development of remote sensing cloud computing platform in China and the application of domestic satellite data in cloud computing platform.

Key words: remote sensing, big data, remote sensing cloud computing platform, distributed storage, parallel computing