第2卷 第2期 2020年4月
Vol.2 No.2 Apr.2020

ISSN 2096-742X
CN 10-1649/TP

文献DOI：
10.11871/jfdc.issn.
2096-742X.2020.
02.003

文献PID：
21.86101.2/jfdc.
2096-742X.2020.
02.003

页码：31-39

开放科学标识码
（OSID）

IA：一种科学数据云分析服务管理引擎

孟珍[1,2]，王学志[1,2]，谢志敏[3]，胡良霖[1,2]，陈之端[2,4]，马俊才[2,5]，佟继周[2,6]，张艳玲[7*]，周园春[1,2*]

1. 中国科学院计算机网络信息中心，北京 100190
2. 中国科学院大学，北京 100049
3. 海参军事海洋环境建设办公室，北京 100081
4. 中国科学院植物研究所，北京 100093
5. 中国科学院微生物研究所，北京 100101
6. 中国科学院国家空间科学中心，北京 100190
7. 中国烟草总公司郑州烟草研究院，河南 郑州 450001

摘　要：【目的】随着科学大数据技术的发展，问题导向的数据端分析成为常态。科学数据处理以云计算的形式跑在数据端，并提供安全的用户访问方式、可选的算法资源库、高效的数据存取接口、便捷的用户交互工具、有效扩展的计算和存储资源，将有力提升科学家的数据分析探索效率。【方法】本文提出一种基于容器技术的科学数据端云分析服务管理引擎设计方案：资源节点以自动注册的方式进行横向扩展，资源节点可以是物理主机或虚拟主机；当在用资源达到阈值，管理节点通过接口启动资源节点的注册，同时资源入池；可选的算法资源库、高效的数据和计算访问接口均以容器镜像的方式进行版本控制，在构造资源池时选用。容器实例池的健康度在节点内部进行维护，根据用户的最长使用时间、静默时间等进行实例生命周期管理；内部资源池的容器实例有准备中、准备好、使用中、消亡中几种状态，并始终维护资源池的固定大小。用户认证访问时，根据用户的领域算法库的选择和资源池的使用率进行新用户资源的接入，并通过代理配置提供唯一的标识入口以供用户访问；用户以安全加密的网络访问方式访问交互编程组件或交互应用组件，即可使用数据端的数据资源和计算资源。每个交互组件均在独立的容器实例中，可以进行有效的资源隔离。【结果】基于以上科学数据端云分析服务管理引擎构建的交互分析云服务系统 IA（Interactive Analysis Cloud Service System）V1.0，实现了科学数据端云分析资源的统一管理服务，可以通过服务门户直接面向终端科学家使用，也可以通过 API 接口以 docker 容器交付的方式给其它现有数据系统调用。已逐步构建生命健康、生态环境、气象水文等领域的科学数据端云分析服务，已应用于中国科学院战略先导专项 A、中国科学院战略先导专项 B、国家烟草专卖局重大专项等重大项目；已应用于国家微生物科学数据中心、国家空间科学数据中心等国家科学数据中心；已应用于地理空间数据云、DarwinTree 分子数据与应用环境等领域公共平台。并提供面向 R、TensorFlow、Data Science、All Spark 等的常用工具服务，用户可以 https 的方式访问交互编程组件（iJupyter）或交互应用组件（iWorkflow），即可使用数据端的数据资源和计算资源。

关键词： 数据端分析；大数据；容器技术；IA；云服务及管理

基金项目：中国科学院战略先导专项（XDB31000000）；中国烟草总公司科技重大专项（110201901025（SJ-04））；广东省生物医药计算重点实验室（2016B030301007）；中国科学院海洋大科学研究中心重点部署项目（COMS2019Q17）；中国科学院“十三五”信息化建设专项（XXH13504-03；XXH13506-102）

*通讯作者：周园春（E-mail: zyc@cnic.cn）；张艳玲（E-mail: zhangyanling@ztri.com.cn）

IA: An Interactive Analysis Service Management Engine in Scientific Data Cloud

Meng Zhen [1,2], Wang Xuezhi [1,2], Xie Zhimin [3], Hu Lianglin [1,2], Chen Zhiduan [2,4], Ma Juncai [2,5], Tong Jizhou [2,6], Zhang Yanling [7*], Zhou Yuanchun [1,2*]

1. Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China
2. University of Chinese Academy of Sciences, Beijing 100049, China
3. Naval Military Marine Environment Construction Office，Beijing 100081, China
4. Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
5. Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China
6. National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China
7. Zhengzhou Tobacco Research Institute of CNTC, Zhengzhou, Henan 450001, China

Abstract: **[Objective]** With the development of scientific big data technology, problem-oriented analysis becomes normal case. Therefore, in views of the high cost of data migration and the reliance of data analysis on scientific big data, it is necessary to provide a scientific data analysis service engine in the data cloud, providing efficient extended computing and storage resources, optional algorithm resource libraries, and high-efficiency access interfaces with convenient user interaction tools and secure user access policy. Then, scientists can get rid of problems including large-scale data migration and adaptation to programming languages, algorithm environments, version issues, and resource calls, etc.. **[Methods]** An interactive analysis service management engine in scientific data cloud is presented. In our solution, resource nodes are scaled out through automatic registration. Resource nodes can be physical hosts or virtual hosts. When the utilization rate of computing resources reaches the threshold, the management node starts resource registration. Subsequently, a resource host is to be registered and the available container instances are added into the pool. The optional algorithm resource libraries, high-efficiency access interfaces for data resources and computing resources are versioned in the form of container mirrors for constructing the computing resource pools. The health of the container instance pool is maintained inside of the host. The instance lifecycle management is performed according to the maximum usage time and maximum silent time of each instance. With the always maintained fixed size resource pool, the container instance of the internal resource pool is in one out of four states, that is, preparing, ready, in use, and disappearing. There are several components set in the scientific analysis service system, including the proxy component, the orchestration module component, the user authentication component, the monitoring management component, buffer component, and a cache database. When a user accesses, the resources are conveyed according to the algorithm library selection and resource pool utilization rate, and a unique identity port (PID) is assigned for user access through proxy configuration. The access is in a secure encrypted network to interact with programming components or interactive application components that can use data and computing resources on the cloud. Each interactive component is in a separate container instance for effective resource isolation. **[Results]** Based on the interactive analysis service management engine in scientific data cloud, iAnalysis (IA for short), an interactive analysis cloud service system V1.0, gives a unified cloud resource management service for scientific data analysis. It can not only be used directly by end-user scientists through the IA's service portal, but also be called by other existing data systems in the form of docker container. By now, IA has provided several scientific cloud analysis services in the fields of life and health, ecological environment, meteorology, and hydrology, etc. It has been applied to major projects such as the Strategic Priority Research Program of the Chinese Academy of Sciences (both A and B) and the Major Project of the State Tobacco Monopoly Administration. It has also been applied to several National Scientific Data Centers, such as the National Microbial Science Data Center, National Space Science Data Center, and public platforms such as GSCloud (www.gscloud.cn) and DarwinTree (www.darwintree.cn). It also provides common coding tools for "R", "TensorFlow", "Data Science", "All Spark", and so on. Users can access the interactive programming component (iJupyter) or interactive application component (iWorkflow) through https to use data resources and computing resources of the data cloud.

Keywords: analysis in data cloud; big data; container technology; IA; cloud services and management

引言

随着科学大数据技术的发展，问题导向的数据端分析成为常态。一方面在专业科学垂直领域，随着传感器布网的增多、采样指标的扩展和采样频率的密集，数据端的数据量级也极具膨胀，数据迁移的时间和空间成本代价增大、存取效率亟待提高[1-2]；另一方面科学数据挖掘分析所用的计算资源也随着数据规模的扩大越来越多，并且这些数据和资源会随着分析的不同有弹性的需求[3]；再者，领域科学家在进行数据分析处理时，需要多样的算法模型和工具库，由此编程语言、算法环境、适配版本、资源调用等问题也往往是其不得不额外付出精力的多个方面[4-5]。

比如在生命健康领域，有来自国际 HapMap 项目[6]近十亿个验证的基因型，也有通过 dbGap[7]提供的大量公共资助研究的数十亿基因型和基因表达测量等，为涉及人类疾病基因组学方面的翻译研究创造了巨大的机会，但也呈现出一些危机。虽然有许多有用和强大的注释资源、数据管理和分析能够与特定的可用数据一起运行，但它们往往也作为独立的“孤岛”存在；虽然领域正在投入大量资源来积累遗传和基因组数据库并在创建强大的社区支持的软件中进行分析，但是现在可以将所有这些资源有效地纳入日常研究实践中的领域科学家的比例仍然相对较小。部分原因是数据以各自的数据格式存在，并不总是很好地扩展到大数据集合，且不在任何单一的集成框架中。用户在研究工作中，需要将数据从一个数据存储库或应用程序输出转换为另外一个特定格式才能进一步处理。

再如在地学遥感领域，互联网公开了大量观测数据和处理软件，分发免费数据和社区支持的分析软件[8]。获得适当信息技能的人员可以创建新的软件基础设施，以构建可重复的流程来转换数据，并将所有这些丰富的资源整合到其工作中。然而，大多数领域科学家在获得新的重要方法和数据方面受到限制，在基础资源使用方面也受到制约。

由此，科研数据处理以云计算的形式让分析跑在数据端，并提供安全的用户访问方式、可选的算法资源库、高效的数据存取接口、便捷的用户交互工具、有效扩展的计算和存储资源，将有力提升科学家的数据分析探索。

在该背景下，不同的应用项目也正在逐步前行。SciServer[9]是由美国国家科学基金会（NSF）支持的项目，旨在建立一个集成网络基础设施系统，能提供免费的科学数据发布平台，以便从观测和模拟中获取大量数据集。系统提供了从天文学到基因组学等多个 pb 级科学数据集的访问，也提供了一组简单但功能强大的基于浏览器的工具，用于可视化和分析这些数据集。Galaxy[10]是一个开放的基于网络的基因组研究平台，旨在解决随着生命科学对计算方法的日益依赖，计算结果的可访问性和可重复性的担忧。Galaxy 是基于 web 交互式，为用户提供了一种交互完整计算分析的媒介。Galaxy 自动跟踪和管理数据来源，并为捕获计算方法的流程提供支持。

本文基于科学数据分析在数据使用、算法环境、编程语言、适配版本、资源调用等方面的问题，提出一种基于容器技术的科学数据端云分析服务管理引擎设计并进行系统实现，发布交互分析云服务系统 V1.0，应用于多个国家科学数据中心、两个领域公共平台和若干重大项目。以下从 IA 关键技术方法、应用及下一步发展几个方面进行介绍。

1　IA 关键技术方法

本文提出一种基于容器技术的科学数据端云分析服务管理引擎设计，其资源节点以自动注册的方式进行横向扩展。资源节点可以是物理主机或虚拟主机；当在用资源达到阈值，管理节点通过接口启动资源节点的注册，同时资源入池；容器实例池的健康度在节点内部进行维护，根据用户的最长使用时间、静默时间等进行实例生命周期管理；内部资源池的容器实例有准备中、准备好、使用中、消亡

中几种状态，并始终维护资源池的固定大小。用户认证访问时，根据用户的领域算法库的选择和资源池的使用率进行新用户资源的接入，并通过代理配置提供唯一的标识入口以供用户访问；用户以安全加密的网络访问方式访问交互编程组件或交互工作流组件，即可使用数据端的数据资源和计算资源。每个交互组件均在独立的容器实例中，可以进行有效的资源隔离。面向领域科学交互式大数据分析服务的方法流程设计如图 1 所示。

1.1 科学数据端云服务方法

一种基于容器技术的科学数据端云分析服务方法，其步骤为：

（1）创建主机节点池：科学数据端云分析系统中建立一组主机节点池；并初始化一组管理组件和一组容器资源池；所述主机节点池包括一个管理节点和若干资源节点，所述管理节点和资源节点可以是物理主机或虚拟主机；

（2）注册资源节点：当资源利用达到阈值时，管理器启动资源节点的注册并记录相关参数到缓存器；当任一项资源利用率的超阈值计数次数达到该项资源利用率的超阈值计数次数阈值，管理器启动主机节点池中的资源节点进入备用状态并记录资源节点的创建时间、节点地址参数到缓存器；

（3）创建容器实例池片：启动注册的资源节点以分析算法库为区分建立面向相应分析算法库的容器实例池片，创建并启动容器实例，并启动每个容器实例内的服务，记录容器实例池片相关参数和每个容器实例相关参数（即容器实例信息）到缓存器；并配置实例代理，代理器通过代理配置提供容器实例 Web 访问唯一标识供用户访问。

（4）启动维护器对容器实例池片的维护：维护器定期读取缓存器中该维护器所在的资源节点内部的容器实例池片相关参数进行维护。维护器定期读取缓存器中的消亡列表中的待消亡容器实例个数和待消亡容器实例的实例名；删除以上列表中每个待消亡容器实例；删除缓存器中的以上步骤的消亡列表中的待消亡容器实例的实例名；每删除一个容器实例，随后启动一个容器实例的创建并追加入缓存器中可用容器实例列表。

（5）接入用户服务：根据用户对分析算法库的选择，择优选择对应容器实例；同时判断是否需要加入新的后备容器实例池片入容器实例池，如果需要，则执行步骤 B。

1.2 IA 服务管理引擎设计

IA 服务管理引擎的管理组件包含：缓存器、代理器、接入器、管理器、监控器、镜像仓库和维护器。IA 管理引擎架构设计如图 2 所示，其中：缓存器、代理器、接入器、管理器、监控器、镜像仓库运行于管理节点上，维护器运行在资源节点上。

管理器进行资源节点的注册和容器实例池的注册，并可以修改容器实例池和容器实例的参数阈值并更新到缓存器。

监控器启动定时任务，监控资源节点和节点上每个容器实例的性能指标写入缓存器。

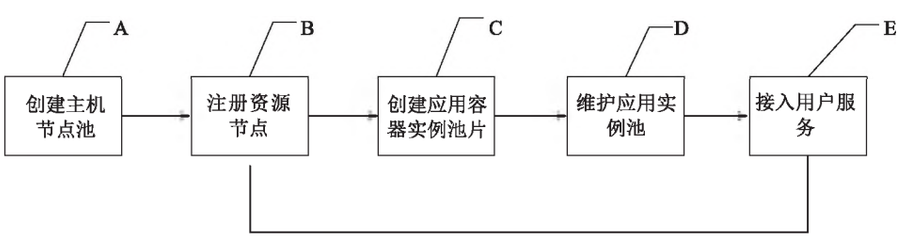

图 1 IA 方法流程设计

Fig.1 IA method flow design

代理器配置代理每个容器实例以提供容器实例Web访问唯一标识，并获取用户的使用情况。

接入器计算最优接入实例池片，并返回该最优实例池片的一个容器实例的唯一标识以供用户访问；所述最优的计算方式是对用户选择容器资源池的各个容器实例利用率进行计算，选择容器实例利用率最低的进行接入。

缓存器存储相关的资源节点信息、容器实例池信息、容器实例池片信息、容器实例信息。

所述的资源节点信息至少包括主机地址和定时监测的性能指标等数据；所述的容器实例池信息至少包括容器实例池标识、算法库模板名称标识、容器实例池所含容器池片标识等数据；所述的容器实例池片信息至少包括该容器实例池片名称标识，该容器实例池片容量阈值，该容器实例池片所含容器实例名称标识、创建时刻、最新维护时刻等数据；所述的容器实例信息至少包括该容器实例名称标识、容器实例参数（容器实例的创建时刻、最新接入时刻、最新活动时刻、最大存活时长、最大静默时长）和定时监测的性能等数据。

维护器定时从缓存器中读取该资源节点的容器池片相关信息、容器实例相关信息、容器实例消亡列表等，在该资源节点内部进行维护以对进入容器实例消亡列表的容器实例进行停止和删除，并进行新容器实例的生成和入容器实例池。

镜像仓库存储以分析算法库为区分的分析算法库镜像，以供资源节点上每个容器实例池片的创建和容器实例池片的内部实例维护。

1.3 IA 容器资源拓扑

在IA服务管理引擎中，容器资源池由若干以分析算法库为区分的容器实例池组成，即每一容器资源池包括多个容器实例池；每一分析算法库对应一个容器实例池；所述容器实例池由分布在不同资源节点上的以相同的分析算法库镜像产生的容器实例

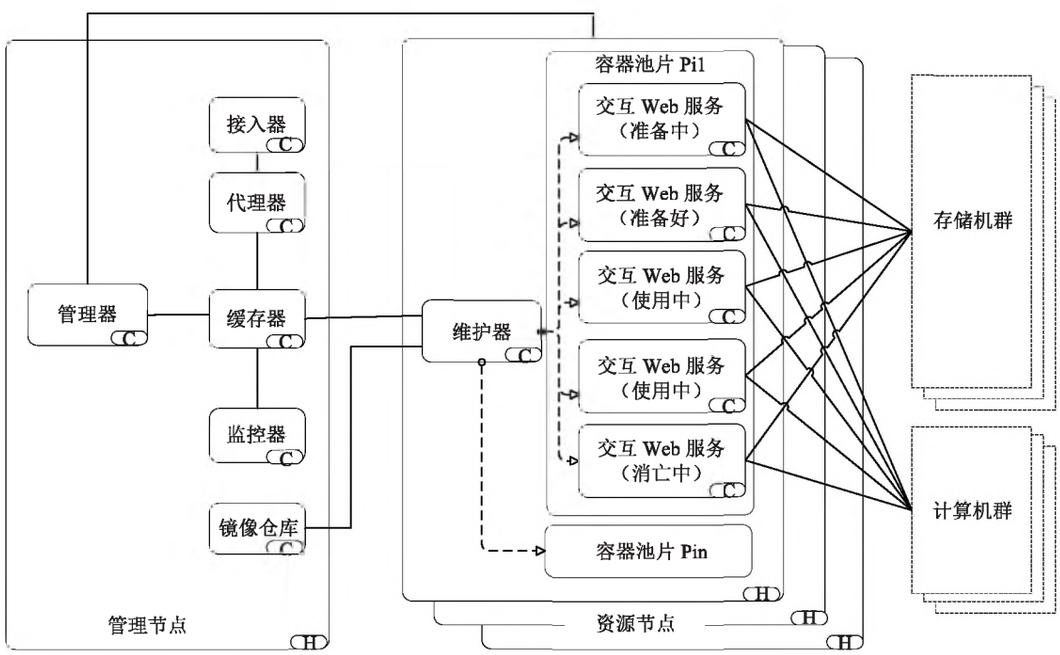

图 2 IA 管理引擎架构设计

Fig.2 IA platform architecture

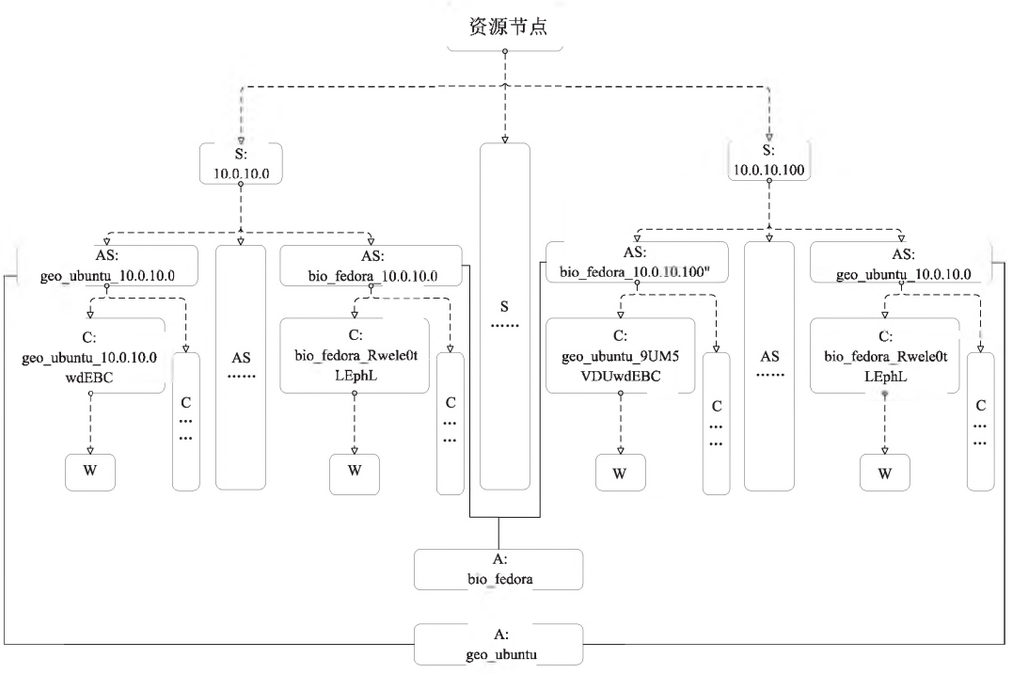

图 3 IA 资源拓扑图

（资源节点-容器实例池-容器实例池片-容器实例-应用服务程序）

Fig.3 Resource topology diagram for IA

(resource node-container instance pool-container instance pool slice-container instance-application server)

池片组成；容器实例池片由一定数量的以相同的分析算法库镜像产生的容器实例组成；容器实例内部均有一个科学数据端服务；容器实例是共享宿主机的 CPU、内存、存储等。

所述每个资源节点可以支持建立多个分析算法库的应用容器实例池片，即同一资源节点可以作为不同容器实例池中的容器实例池片。 资源节点、容器实例池、容器实例池片、容器实例和面向科学分析的 Web 应用，其拓扑关系如图 3 所示。每个资源节点（S）包括若干容器资源池片（AS）；每个以分析算法库区分的容器资源池（A）的容器资源池片可以分布在不同的资源节点上；每个容器实例（C）内含一个面向科学分析的 Web 应用（W）。

2 IA 平台的应用

目前，基于以上科学数据端云服务管理引擎已构建 IA：交互分析云服务系统 V1.0，实现科学数据端云分析资源的统一管理服务，可以通过 IA 的服务门户进行直接面向终端科学家试用，也可以通过 API 接口以 docker 容器交付的方式给其它现有数据系统调用。系统支持中英文操作界面如图 4-5 所示。

图 4 IA：交互分析云服务系统 V1.0（中文版）

Fig.4 IA: Interactive Analysis Cloud Service System V1.0 (Chinese)

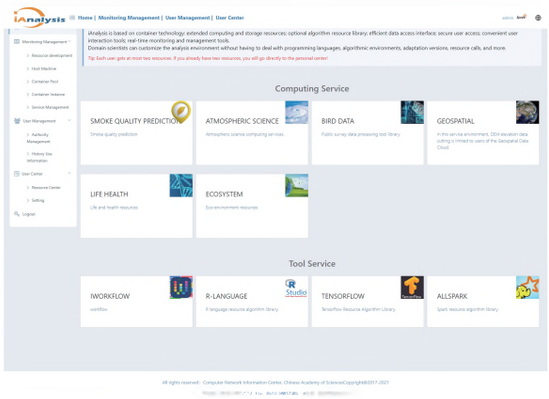

图 5 IA：交互分析云服务系统 V1.0（英文版）

Fig.5 IA: Interactive Analysis Cloud Service System V1.0 (English)

IA：交互分析云服务系统 V1.0 可以基于自有的基础设施独立部署，也可以通过构建在中国科学院计算机网络信息中心基础设施之上的云服务版本进行资源试用。此云服务版本提供 R(https://www.r-project.org/)、TensorFlow[11]、Data Science、All Spark[12] 等的常用工具服务，并面向领域提供定制服务，已逐步构建生命健康、生态环境、大气科学、烟草行业等领域的基础算法库。用户可以 https 的方式访问交互编程组件（iJupyter）或交互应用组件（iWorkflow），即可使用数据端的数据资源和计算资源。

2.1 IA 应用于重大项目

IA：交互分析云服务系统 V1.0 以实际应用需求为牵引，逐渐为众多重大项目提供工具和分析支撑服务：面向中国科学院战略先导专项 A“地球大数据工程”的“地面监测数据资源汇聚和共享”，提供公众调查类数据处理工具库服务鸟类数据的汇聚处理；面向国家烟草专卖局重大专项“烟草科研大数据”的“烟叶质量大数据分析挖掘技术及应用研究”，提供烟叶质量工具库服务烟叶质量的模型训练与预测；面向中国科学院战略先导专项 B“大尺度区域生物多样性格局与生命策略”的“亚热带森林群落多样性格局与生命策略”课题，提供系统进化分析工具库服务中国生命之树构建和生物多样性制图等，并同时面向气象水文、海洋分析、代谢组学分析等应用场景进行应用。

2.2 IA 应用于国家科学数据中心

IA：交互分析云服务系统 V1.0 应用于国家微生物科学数据中心，面向微生物领域云等 6 大类 20 多个数据源、1 716 272 390 数据实体提供 Metagenome Tools、Alignment、Evolution Analysis 等 90 余个模型分析服务。

在 IA 的功能基础上，定制形成 iSpace 并成功应用于国家空间科学数据中心。iSpace 打通了内部 iWorkflow 和 iJupyter 的交互通道，既能在 iWorkflow 里分析数据、构建工作流，同时能利用 iJupyter 对工作流中的工具模块进行编辑。构建 Tsyganenko 磁场模型库、地磁截止刚度计算等若干领域专业工具 / 模式库，和支持户自定义工具封装与发布。

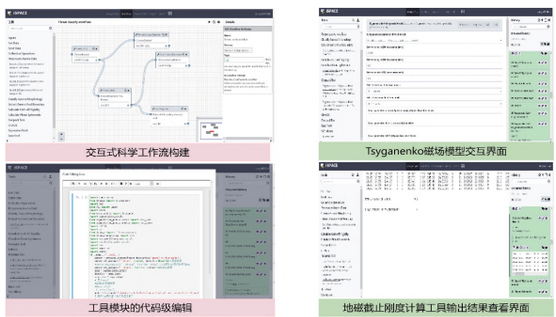

图 6 基于 IA 的 iSpace（国家空间科学数据中心）

Fig.6 iSpace based on IA

2.3 IA 应用于领域公共平台

IA 通过 API 面向 GSCloud 地理空间数据云[13] 提供应用服务，支持地理空间数据云 9 大类 105 种数据资源、600TB 地理空间数据资源使用，支持遥感数据切割、地图服务发布、植被指数模型运算等几大类算法库，面向 34 万注册用户提供科学数据端的在线交互分析。用户以 https 的方式访问交互编程组件（iJupyter）即可以使用。

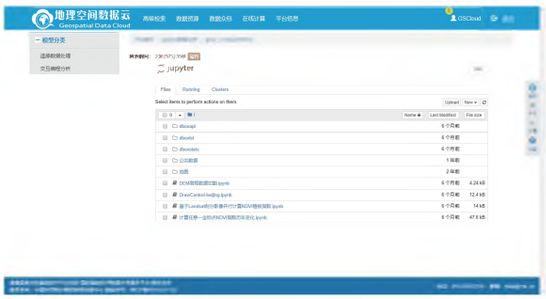

图 7 基于 IA 的在线交互编程分析（GSCloud）

Fig.7 Interactive programming analysis based on IA

IA 通过 iWorkflow 面向 DarwinTree 分子数据与应用环境[14]，同步标记的序列数据 200 007 147，覆盖中国领域 3114 属级单元，提供 130 余模型服务，支持中国维管植物系统发育大树构建和挖掘、中国被子植物群的进化史分析等研究，支持研究团队相关成果发表在《自然》获得 2018“中国生命科学十大进展”[15]。

3 展望与下一步工作

科学数据的应用领域在不断扩展，科学数据端分析需求也在不断发展。紧跟科学领域的发展需要，构建针对不同应用场景的算法库支持多样的编程语言、算法环境和适配版本的资源池，并对具体性能问题结合领域方法进行调优，不断丰富科学数据端的计算资源，将是一个不断和领域科学家协同发展的过程。IA 将继续以中国科学院重大专项和行业科学领域需求为牵引，以 GSCloud、DarwinTree、中国科学数据云等平台服务为依托，进一步完善面向地理空间、系统进化、国家基础学科公共科学数据中心等的公共服务，并针对空间科学大数据、海洋科学大数据、烟草行业等特色领域需求进一步进行丰富发展，夯实打磨 IA 的服务场景和服务能力。

利益冲突声明

所有作者声明不存在利益冲突关系。

参考文献

[1] Vanderbilt, K., & Gaiser, E. The international long term ecological research network: a platform for collaboration. Ecosphere,2017, 8(2).

[2] Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., ... & Fuentes, J. FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities[J]. Bulletin of the American Meteorological Society, 2001,82(11): 2415-2434.

[3] Lee, J. G. & Kang, M. Geospatial big data: challenges and opportunities[J]. Big Data Research, 2015,2(2):74-81.

[4] Labrinidis, A., & Jagadish, H. V. Challenges and opportunities with big data[J]. Proceedings of the VLDB Endowment, 2012,5(12): 2032-2033.

[5] Merelli, I., Pérez-Sánchez, H., Gesing, S. & D'Agostino, D.Managing, analysing, and integrating big data in medical bioinformatics: open problems and future perspectives[C]. BioMed research international, 2014.

[6] Gibbs R A, Belmont J W, Hardenbol P, et al. The International HapMap Project[J]. Nature, 2003, 426(6968): 789-796.

[7] Tryka K A, Hao L, Sturcke A, et al. NCBI's Database of Genotypes and Phenotypes: dbGaP[J]. Nucleic Acids Research, 2014: 975-979.

[8] Woodcock C E, Allen R G, Anderson M C, et al. Free access to Landsat imagery[J]. Science, 2008, 320(5879): 1011-1011.

[9] Medvedev D, Lemson G, Rippin M, et al. SciServer Compute: Bringing Analysis Close to the Data[C]. statistical and scientific database management, 2016.

[10] Blankenberg D, Von Kuster G, Coraor N, et al. Galaxy: a web-based genome analysis tool for experimentalists. [J]. Current protocols in molecular biology, 2010, 89(1).

[11] Abadi M, Barham P, Chen J, et al. TensorFlow: a system for large-scale machine learning[J]. 2016.

[12] Zaharia M, Chowdhury M, Franklin M J, et al. Spark: cluster computing with working sets[C]// Usenix Conference on Hot Topics in Cloud Computing. USENIX Association, 2010:10-10.

[13] Xuezhi W, Jianghua Z, Yuanchun Z, et al. The Geospatial Data Cloud: An Implementation of Applying Cloud Computing in Geosciences[J]. Data Science Journal, 2014, 13:254-264.

[14] Meng Z, Dong H, Li J, et al. Darwintree: A Molecular Data Analysis and Application Environment for Phylogenetic Study[J]. Data Science Journal, 2015.

[15] Lu L M, Mao L F, Yang T, et al. Evolutionary history of the angiosperm flora of China[J]. Nature, 2018.

收稿日期：2020 年 1 月 7 日

周园春，中国科学院计算机网络信息中心，博士，研究员，博士生导师，中国科学院特聘研究员，中心主任助理，中心学位评定委员会主席，大数据技术与应用发展部主任，大数据分析与计算技术国家地方联合工程实验室秘书长，国家烟草专卖局烟草科研大数据重大专项技术首席。发表 SCI/EI 收录论文 90 多篇。主要研究方向为云计算、大数据分析与处理。

本文主要承担工作为 IA 整体架构设计。

Zhou Yuanchun is the research fellow, Ph.D. supervisor and the assistant director in Computer Network Information Center, Chinese Academy of Sciences and the director of the Department of Big Data Technology and Application Development. He is also the chairman of the Degree Evaluation Committee in Computer Network Information Center, Chinese Academy of Sciences. His research interests include cloud computing, big data analysis and processing. He has published more than 90 papers included in SCI/EI.

In this paper he is mainly responsible for the overall framework design of IA.

E-mail: zyc@cnic.cn

孟珍，中国科学院计算机网络信息中心，高级工程师，硕士研究生导师，大数据技术与应用发展部数据资源与应用实验室副主任，主要研究方向为多源异构数据的融合管理与关联技术、面向领域大数据分析模型与云服务技术。发表 SCI/EI 收录论文 20 多篇。本文主要承担 IA 方法研究及应用。

Meng Zhen is a senior engineer and the master supervisor at the Department of Big Data Technology and Application Development at Computer Network Information Center, Chinese Academy of Sciences. She is the deputy director of the Resource and Application Lab at the Department of Big Data Technology and Application Development. Her research interests include big data management, processing, mining, analysis and other related technologies. And she has published over 20 papers included in SCI/EI.

In this paper she is mainly responsible for methods research and platform overview of IA.

E-mail: zhenm99@cnic.cn

引文格式：孟珍,王学志,谢志敏,等. IA：一种科学数据云分析服务管理引擎[J]. 数据与计算发展前沿,2020,2(2):31-39.DOI:10.11871/jfdc.issn.2096-742X.2020.02.003.PID:21.86101.2/jfdc.2096-742X.2020.02.003.

Meng Zhen,Wang Xuezhi, Xie Zhimin, et al.. IA: An Interactive Analysis Service Management Engine in Scientific Data Cloud[J].Frontiers of Data & Coputing,2020,2(2): 31-39.DOI:10.11871/jfdc.issn.2096-742X.2020.02.003.PID:21.86101.2/jfdc.2096-742X.2020.02.003.