论 文

天宫二号应用任务地面数据处理与服务系统

李盛阳^{1,2*}, 刘志文^{1,2}, 张万峰^{1,2}, 谭洪^{1,2}, 刘云飞^{1,2}, 王博^{1,2}, 顾逸东¹

1. 中国科学院空间应用工程与技术中心, 北京 100094;

2. 中国科学院太空应用重点实验室, 北京 100094

* E-mail: shyli@csu.ac.cn

收稿日期: 2019-09-18; 接受日期: 2019-11-14; 网络版发表日期: 2020-05-13

国家重大科技专项(编号: Y3140231WN)资助项目

摘要 面向天宫二号应用任务各类不同机理的先进科学实验类载荷和对地观测载荷的数据处理、存档管理、质量分析、运营服务和数据应用研究等方面的需求, 研制建设了一套业务功能覆盖全面、支持能力突出、技术特色鲜明的一体化地面数据处理与服务系统, 实现了多源海量数据自动化并行处理、高精度标准化产品生产、高效数据存档与可视化管理、数据质量分析与控制、在线分发服务及运营等功能。本文总结了国内外航天地面数据处理与服务系统现状及发展趋势, 分析了天宫二号应用任务空间科学与应用数据在处理、管理、共享服务等方面的技术要求, 系统性地设计了地面数据系统的整体架构、业务流程、功能与接口等, 详细设计并实现了数据并行处理与产品生产、可视化数据存储与管理、系统运行管理与控制、数据质量分析与控制及数据在线共享服务等方面的业务功能, 进一步分析了新型对地观测载荷高精度数据处理、业务流程高并发调度与管控技术等关键技术问题并阐述了具体解决方案。该系统已稳定运行三年, 为国内外广大科研工作者累计提供了超过48 TB的数据产品, 为中国载人航天工程空间应用各领域科学研究、成果产出及应用推广发挥了重要的地面数据系统技术支撑与保障作用。

关键词 天宫二号, 地面数据系统, 数据处理与服务, 数据分发共享

1 引言

天宫二号作为我国第一个真正意义上的太空实验室, 在轨期间开展了大量空间科学实验和技术试验, 包括地球科学观测及应用、空间科学实验及探测、应用新技术等领域的十余项高精尖的任务, 其中地球科学观测及应用包括宽波段成像仪、三维成像微波高度计和多波段紫外临边成像光谱仪等3类体制各异的载荷。

空间应用任务中存在的主要问题和挑战有以下几方面: (1) 载荷机理新、技术先进、实验差异大, 导致地面数据处理算法复杂多样, 创新性要求高; (2) 在轨高速下行的数据对地面数据处理的时效性要求高; (3) 应用研究领域差异大, 不同类别/级别数据产品需求多样, 应用研究的数据产品精度要求高; (4) 科学实验和地球观测的长期运行对及时掌握数据变化情况, 开展数据产品质量分析与评价的需求迫切; (5) 数据类型多、

引用格式: 李盛阳, 刘志文, 张万峰, 等. 天宫二号应用任务地面数据处理与服务系统. 中国科学: 技术科学, 2020, 50: 1066–1080

Li S Y, Liu Z W, Zhang W F, et al. Tiangong-2 application mission ground data processing and service system (in Chinese). Sci Sin Tech, 2020, 50: 1066–1080, doi: 10.1360/SST-2019-0331

体量大,造成数据存档、管理难度大;(6)大范围数据在线服务的便捷性、时效性要求高等。作为天宫二号应用任务的重要研究内容之一,亟需研制建设一套功能完善、支持多任务的天宫二号应用任务地面数据处理与服务技术系统,为各领域科学研究与应用、数据成果产出等提供全面的地面数据系统技术支撑与保障。

国内外科研与应用机构在航天任务地面数据系统建设方面,综合运用分布式计算、海量数据库存储技术、高速通信技术等现代化信息技术,取得了卓有成效的进步与发展,研制建设出了满足各类航天任务应用需求的地面数据系统,其中有代表性的有:美国国家航空航天局(NASA)建立的地球观测系统数据与信息系系统(Earth observing system data and information system, EOSDIS)^[1]负责所有地球科学观测数据进行统一的存档、管理和分发,同时对部分载荷数据经过标准化产品生产后再发布。该系统采用开放式分布式系统架构,按照地球科学学科划分成多个分布式数据存档中心(distributed active archive center, DAAC),每个 DAAC 负责一个特定学科的数据产品的生产、存档、管理及分发,采用统一的扩展分层格式(HDF-EOS)高效存储相关数据,并为系统中的数据产品提供统一的访问接口,通过FTP和Web等形式提供数据分发共享服务。欧洲空间天文学中心(European Space Astronomy Centre, ESAC)地面数据系统^[2]负责对太空望远镜获取的大量宇宙天文数据与信息进行存储、处理与共享等,该系统利用数据存储层、业务逻辑层、交互层等模块化3层架构,数据存储和数据展示相互隔离,有助于后期的维护升级。数据存储层负责数据和元数据等存储管理,数据存储在UNIX文件系统,元数据存储在SYBASE关系数据库中,数据导入和元数据提取单独进行,方便添加新数据及应对需求的变化;业务逻辑层提供对数据产品和元数据的透明访问,部署在存档服务器上;交互层通过图形用户界面或虚拟天文台界面为用户提供数据共享服务。

国内地面数据系统主要有国家卫星气象中心、中国资源卫星应用中心等单位建立的地面数据系统。国家卫星气象中心的数据存储与服务系统^[3]由多个高性能计算机集群组成,应用可配置的动态加载容器和多机作业并发处理技术将集群构建为高扩展性、高可用、高效率的业务系统,主要包括数据存档与管理、

数据检索与订购、空间数据库和WebGIS发布、运行监控和统计、用户服务和支持等分系统,共同构建了支撑风云卫星系列任务的数据存档与服务体系。中国资源卫星应用中心的地面数据系统(<http://www.cresda.com/CN/fwyytyx/dmxtsjcl/4417.shtml>)采用集中存储、系统管理和分布式处理的体系结构,数据共享服务采用Web方式提供数据检索访问等,利用多节点并行计算、计算节点在本地磁盘构建文件系统等计算机技术,结合遥感卫星数据特点,大幅提升资源系列遥感卫星的地面系统的计算和存储能力。

综上所述,现有地面系统特点及发展趋势如下。

(1) 多任务、多平台、多载荷的地面数据系统。随着空间应用任务的不断推进,地面数据处理与服务系统朝着处理与管理多任务、多平台、多载荷的空间科学数据的趋势发展,而不仅仅局限于单一任务、单一平台、单一载荷的地面数据处理与服务系统。

(2) 清晰可扩展的分层架构体系。随着数据采集手段增强及在轨运行时间的增加,采集的数据量急剧增长,地面系统采用分布式结构存档海量数据,利于后续数据存储空间的扩容与拓展,同时采用模块化灵活性高的系统架构,对数据存储和数据展示进行隔离等,利于后期系统的维护与升级。

(3) 高效的数据处理、存储、分发能力。采用高性能并行计算技术和工作流程管理技术等,通过高性能的作业调度和并发 workflow 管理,实现对海量、多源、异构数据快速高效地处理、存储、分发等综合功能。

面向天宫二号应用任务的实际需求,结合地面数据系统技术发展情况,研制建设了一套从数据并行处理、数据归档管理、数据产品可视化、数据质量分析与控制到数据分发共享服务等一体化的高性能地面数据处理与服务系统,该系统具有实时调度的高并发系统管控、基于CPU+GPU数据并行处理、数据快速校正处理、数据高精度处理与产品定制化生产、多载荷数据快速可视化归档与管理、基于三维地球的数据产品可视化、体系化的数据产品质量分析与评价等特色,创新性提出了虚拟相机模型的宽波段图像拼接、基于内陆水体区域的干涉系统参数定标、紫外临边成像光谱仪切点高度定位、业务流程高并发调度与管控等4类方法。在天宫二号应用任务及天舟一号货运飞船的各类数据应用及推广工作中起到了重要的系统支持作用。

2 地面数据系统技术要求分析

2.1 数据处理精度高

天宫二号空间实验室开展包括地球科学观测及应用等十余项空间科学试验/实验, 载荷机理新, 技术先进, 实验机理差异大, 涉及数据类型多, 数据处理算法复杂, 需要根据各个实/试验类型, 在进行针对性的解帧, 解格式等预处理外, 地球科学观测及应用数据还需进行标准/定制产品生产, 且处理精度要求高, 如宽波段数据相对定位误差 ≤ 1 个像元, 三维成像微波高度计海上测高误差 ≤ 28 cm, 多波段紫外临边成像数据辐亮度校正误差 $\leq 3\%$ 等。

2.2 数据处理时效性强

天宫二号空间科学与应用数据每天下行的原始数据量超过100 GB, 实验载荷数据类型多样、特征各异、流程差异大。要充分利用工作流控制与并行调度技术, 快速对各类应用载荷的所有下行数据开展集群化、并行化的统一解格式等预处理、空间地球科学领域标准及定制数据产品生产、数据归档存储等, 避免造成每天数据的大量堆积。

2.3 开展数据质量分析与控制

天宫二号应用任务地面数据处理与服务系统要为本领域应用科研用户提供高质量的数据产品, 需开展数据质量分析, 实现对标准数据产品的自动化质量分析、评价, 对高级数据产品的定制分析等。

针对3类对地观测载荷的成像性能会随着长周期在轨运行发生衰减的现象, 需研制定标资料处理软件, 对定期开展的试验场定标试验获取的定标资料进行处理, 优选定标系数, 保障数据产品质量。

2.4 海量多源异构数据要实现高效组织管理

天宫二号应用任务十余项空间科学实验获取的原始数据、各类型各级别数据产品、中间过程数据等数据量级超过200 TB, 数据来源、内容、形式、结构及格式等特征差异大, 传统数据存档与管理方式在海量结构化和非结构化数据存储、海量数据快速高效访问、多源数据整合与管理等方面能力不足, 要制定最优化的数据产品存储策略, 建立合理的数据存储、备份机制与策略, 设计高效的数据存储管理系统, 实现

数据、产品与信息的高效管理。

2.5 数据在线实现高效分发与共享服务

天宫二号空间应用任务数据服务用户众多, 分布世界各地, 并且对数据需求迫切、需求量大且类型各异。为了给用户提供高效的数据在线服务, 根据预测, 需要建设访问并发量 ≥ 1000 人、并发下载线程 ≥ 100 个的数据在线分发共享平台。该平台支持用户身份认证与权限控制操作, 面向任务、公益、公众等多类用户群体, 能够提供内容检索、关联检索、模糊检索等多种数据查询检索、数据浏览和下载, 以及高并发的数据在线分发共享等服务, 满足用户对空间科学数据的应用研究需求。

3 地面数据系统架构与流程设计

天宫二号应用任务地面数据处理与服务系统基于面向服务架构(service-oriented architecture, SOA)设计, 具有松耦合、强扩展、接口协议通用等特点, 支持了宽波段成像仪、三维成像微波高度计、多波段紫外临边成像仪等载荷的数据处理与服务业务, 研究实现了面向各类载荷的辐射校正、几何校正算法, 实现了高精度数据产品生产, 并以标准化服务组件的形式发布数据处理、存档管理、质量分析等服务, 通过工作流引擎高效地组织管理系统业务流程, 基于集群环境利用PBS+MAUI的并行调度算法实现计算节点资源的最大化^[4,5]。

研制的载人航天空间应用数据推广服务平台(<http://www.msadc.cn>), 基于HTTP与FTP等协议, 面向各领域用户提供高精度数据产品的在线数据分发共享服务。

3.1 系统架构设计

地面数据系统硬件架构基于千兆以太网、IB计算网络的双网络, 将高性能Cluster集群、GPU服务器等异构计算资源构建成为统一的计算资源池, 通过管理、调度节点上部署集群调度软件与负载均衡软件确保计算资源的高效利用, 利用并行文件存储系统实现多载荷数据、多业务流程的并发读写, 提高数据并行读写效率。地面数据系统硬件基础架构设计如图1所示。

地面数据系统基础软件架构面向下行数据量大、

图 2 (网络版彩图)地面数据系统基础软件架构设计

Figure 2 (Color online) The design of software infrastructure of ground data system.

数据驱动的自动化数据处理与服务流程由下行原始数据自动触发系统运行管理与控制的流程控制与调度,完成数据处理与标准化产品生产、数据产品存档、元信息入库等功能,全流程实现数据自动化处理与计算/存储资源的动态分配与调度。

订单驱动的定制化数据产品生产与服务流程由用户对数据产品的生产需求转化而来,生成数据定制化生产订单,工作流调度与控制模块对数据产品生产需求进行分类与资源参数解析,读取多载荷数据产品生产流程所需参数信息,提供软件算法参数与数据资源参数,并调度在线计算与数据资源,完成各类型各级别定制化的数据产品生产流程。

数据质量分析与控制流程由系统操作人员发起,采用抽检方式对存档的各类载荷不同级别数据产品进行质量分析与评价,调用不同载荷数据质量分析与评价算法开展评价指标计算,形成数据产品质量分析与评价报告等。

4 系统功能与接口设计

4.1 系统功能

(1) 系统运行管理与控制

在SOA架构基础上将地面数据系统的算法模块管理、资源调度、软/硬件运行状态监控等功能抽象为基础接口,通过标准接口集成多载荷数据处理算法、数据归档管理、质量分析与评价等业务功能,实现多种业务流程的设计与集成、模型与算法的扩展,并在数据处理过程中利用工作流并行调度算法实现系统计算资源的高效利用。

(2) 数据处理与产品生产

面向多类载荷数据产品生产需求如宽波段成像仪、三维成像微波高度计、多波段紫外临边成像仪等,针对各载荷体制各异的成像特征,研究并实现了辐射校正、几何校正等数据产品高精度生产算法。其中,宽波段影像经过暗背景扣除、非均匀性校正、辐射校正、几何精校正等处理;三维成像微波高度计的多视复图像处理涉及原始数据提取、距离压缩、二维成像和多视处理;后向散射系数图像在多视复图像的基础上,经过辐射校正和几何校正^[6]等处理,多视复图像经过干涉处理,生产三维高程图像等;紫外临边数据经过辐照度校正、几何定位、图像压缩等处理。

(3) 数据质量分析与控制

支持宽波段成像仪、三维成像微波高度计、多波段紫外临边成像仪等3类载荷以及不同级别产品的质

量分析与评价, 提供数据质量控制参数, 并实现对标准级别数据产品、定制数据产品的质量分析等。

利用研制的定标资料处理软件, 处理与分析了宽波段成像仪和三维成像微波高度计的试验场定标试验数据, 分别获得了宽波段成像仪各波段的辐射定标系数, 高度计最新的定标常数和系统时延修正参数, 并对数据处理和产品生产中的辐射和几何参数进行了更新, 进而提升数据产品质量。

(4) 数据存档与管理

对处理生成的各级别数据产品进行编目存档, 实现基于不同载荷、不同产品级别、不同采集/生产时间等参数的数据快速查询、检索与高效管理, 各级别数据产品、数据库、元数据等信息数百TB; 提供多载荷数据产品、元数据的远程检索、查询、统计和下载等服务, 并利用磁带库等对原始数据进行备份与恢复, 确保数据的长期有效存储与安全管理。

(5) 数据分发共享

基于在线网络服务架构, 构建面向多任务的数据分发共享服务系统——载人航天空间应用数据共享服务平台(<http://www.msadc.cn>), 通过在线、多维度、多种可视化查询的方式检索可用数据资源, 面向各应用领域用户提供天宫二号、天舟一号货运飞船等空间科学数据产品的查询、下载等分发共享服务。

4.2 系统接口设计

地面数据系统包括数据处理与产品生产、数据归档与管理、系统运行管理与控制、数据质量分析与控制、数据分发共享等5部分业务功能, 各组成部分之间通过WebService协议、Message Queue等消息中间件实现接口调用, 系统运行管理与控制分别与数据处理与产品生产、数据质量分析与控制、数据归档与管理等之间均存在接口, 与数据分发共享之间存在数据传递接口, 同时数据处理与产品生产和数据归档与管理之间存在接口, 如图3所示。

5 系统业务功能设计与实现

5.1 数据处理与产品生产

5.1.1 宽波段成像仪数据处理与产品生产

宽波段成像仪是新一代宽波段、宽视场和“图谱合一”的光学遥感器, 是国内外首次在单台仪器上, 实

现了可见近红外、短波红外和热红外大视场全推扫成像的组合集成功能。宽波段成像仪包括7台相机, 其中3台是可见光近红外波段探测器, 2台是短波红外波段探测器, 2台是热红外波段探测器。这些探测器通过视场拼接组合在一起形成42°观测视场, 在较大范围内获取同一目标的图像和光谱信息。

宽波段影像数据经过辐射校正、几何校正等处理。在轨初期, 辐射校正系数采用实验室定标系数, 该系数利用经过精确标定的积分球作为光源在实验室进行标定获得, 综合标定误差为4.25%, 非均匀性校正精度在0.5%~2%之间; 采用试验场定标资料处理系统开展辐射校正优化。几何校正利用大幅宽虚拟相机模型进行系统级几何校正, 相对误差在1个像素内; 利用参考影像与待校正影像进行自动匹配, 经几何精校正处理, 校正绝对误差在1个像素内。处理结果如图4所示。

5.1.2 三维成像微波高度计数据处理与产品生产

三维成像微波高度计是国际上首个采用近天底角观测和干涉SAR测高技术实现海洋表面和陆地表面宽刈幅、高精度高程数据测量的天基干涉雷达高度计。三维成像微波高度计生产陆地及海洋表面^[7]的多视复图像、后向散射系数图像和三维高程图像。

多视复图像处理涉及原始数据提取、距离压缩、二维成像和多视处理等。后向散射系数图像是在多视复图像的基础上, 经过AGC校正、天线方向图校正、斜距衰减校正、地面分辨单元校正等辐射校正和几何校正^[6]等处理生成; 三维高程图像由微波高度计的通道1和通道2的多视复图像经过干涉处理生成, 干涉处理主要涉及图像配准、干涉图生成、滤波、解缠、基线估计和高程反演等处理^[8]。产品处理精度均满足设计要求。处理结果如图5所示。通过利用先进的多模异构融合方法^[9], 多载荷(宽波段成像仪、三维微波高度计)提供的图像-光谱、高程数据产品, 能够更好地协同服务于天宫二号对地观测数据应用与研究。

5.1.3 紫外临边光谱仪数据处理与产品生产

多波段紫外临边成像光谱仪由紫外环形成像仪和紫外前向光谱仪组成, 首次实现了超大视场对全球中层大气的环形临边紫外辐射特性进行探测, 并采用精细光谱分辨率对前向临边大气的各谱段进行综合探测, 为地球大气环境探测提供新的信息源。

图 3 (网络版彩图)地面数据系统接口设计

Figure 3 (Color online) The interface design of ground data system.

图 4 (网络版彩图)宽波段成像仪数据处理. (a) 原始影像; (b) 辐射校正; (c) 几何校正

Figure 4 (Color online) The wide-band imaging spectrometer data processing. (a) Original image; (b) radiation correction; (c) geometric correction.

紫外临边光谱数据经过辐照度校正、几何定位等处理。辐射校正处理涉及暗信号校正、smear校正和辐

亮度校正等处理; 利用像素视场分布、平台轨道、平台姿态等参数, 将图像每个像素沿视场方向投影到虚拟椭

图 5 三维成像微波高度计图像产品. (a) 成像处理获取的多视复图像; (b) 干涉处理获取的干涉条纹; (c) 后向散射系数图像; (d) 三维高程图像

Figure 5 Image products of the Interferometric Imaging Radar Altimeter. (a) Multi-look complex image; (b) interferometric fringe acquired; (c) backscattering coefficient image; (d) digital elevation model.

球体上, 获得对应像素的目标坐标, 完成几何定位处理. 多波段紫外临边成像光谱数据产品如图6所示.

5.1.4 混合并行数据产品生产

面向宽波段成像仪、三维成像微波高度计等单轨数据成像时间长的特点, 对预处理过程采用并行处理, 实现重叠区可调整的标准景自动划分, 并对标准景数据进行多节点并行化; 面向多波段紫外临边成像仪数据产品生产过程中并行帧数过多的问题, 设计了GPU并行通用模板, 以共享内存方式实现紫外临边数据产品的并行化生产.

对宽波段成像仪、三维成像微波高度计等数据采用MPI并行处理技术, 将此二类数据产品生产过程中的辐射校正、几何校正、产品封装等算法模块标准化, 对其中耗时较长的几何校正过程中的重采样算法进行并行化, 大幅提高数据产品处理效率.

对多波段紫外临边成像仪数据, 以每一轨紫外临边数据纵向排列包含多个标准帧为基本原则, 将紫外临边标准帧数据作为最小单元采用GPU并行处理, 将几何定位算法作为CUDA并行模板中的标准模块, 实现紫外临边数据的多帧并行处理.

通过上述两种并行化方法, 最大化发挥CPU的单

图 6 (网络版彩图)紫外临边成像光谱数据处理. (a) 紫外环形360 nm通道图像产品; (b) 紫外前向光谱仪图像产品

Figure 6 (Color online) The UV imager data processing. (a) Ultraviolet ring imager 360 nm channel image products; (b) ultraviolet forward spectrometer image products.

节点计算优势、GPU的多核高并发计算优势, 实现了基于CPU+GPU混合计算资源的数据产品高效生产, 使数据产品生产日处理能力超过1.2 TB, 支持多领域应用研究的数据需求。

5.2 可视化数据归档与管理

面向待归档的数据集, 可视化数据归档与管理首先调用数据归档流程, 根据数据接口控制文件, 提取数据集元信息, 将关键元信息存入数据库; 然后, 按照数据编目归档规则将数据集归档入分布式文件系统, 形成逻辑完整的数据存档目录树结构。

为了保证数据归档与管理闭环完整并流程可控, 根据不同类型数据特点建立数据存档的可视化管理机制, 设计信息提取和多分辨率可视模型处理算法, 生成数据可视化管理的中间信息数据, 对数据集归档进度和状态进行实时可视化监控和统计分析, 记录并处理归档过程中的异常情况和失败数据, 提高数据管理与服务的效率, 如图7所示, 具体解决方法如下。

(1) 信息提取与信息库构建: 对数据管理过程中涉及的数据覆盖区域、数据统计等进行计算分析, 并将结果存入相关的数据库中。

(2) 数据多分辨率可视基础模型库构建: 由于不同领域的HDF, GeoTiff等标准格式的数据产品精度高、数据量大, 直接对原始产品数据可视化不仅需要集成

各类对象组件, 且占用大量内存, 操作延时明显。为实现数据产品可视化管理与展示以及流畅的交互操作, 对该类标准数据产品中的频谱、影像等数据生成多分辨率高视觉质量的可视模型, 存入分布式数据库。

(3) 金字塔模型及基于瓦片数据并行调度: 为了直观快速地获得空间科学影像数据信息, 使用多分辨率金字塔模型及基于瓦片数据并行调度技术, 通过对影像波段提取与转换, 将空间科学影像数据按空间位置网格划分成大小相等的矩形“瓦片”, 建立多个分辨率递减的瓦片层, 并将其存储到高性能的NoSQL数据库中供访问。设计的多分辨率金字塔及其存储系统, 其逻辑模型如图8所示。

(4) 归档过程可视化监控: 动态、实时地显示各类原始数据和数据产品所处的处理流程和每一流程的处理状态。对数据低质、操作失误、系统故障等因素导致处理失败的数据, 进行自动或人工错误检测、失败分析、消息推送、数据修正或重处理等的处理闭环。

(5) 数据智能检索与管理: 通过文件系统、元信息数据库、多分辨率可视基础模型库和统一数据可视化操作平台对编目存档数据进行管理, 提供对各类数据的智能检索, 支持多个维度数据检索条件的数据与数据产品的关联检索、模糊关键字查询及历史查询条件的智能匹配等, 同时, 根据访问的数据类型、频率等信

图7 (网络版彩图)可视化数据存档与管理示意图

Figure 7 (Color online) The visual data archiving and management diagram.

图 8 (网络版彩图)影像多分辨率金字塔模型

Figure 8 (Color online) The image multiresolution pyramid model.

息, 动态调整元信息搜索索引, 实现智能化的数据查询检索, 提高数据管理、分发共享效率。

5.3 系统运行管理与控制

系统运行管理与控制是天宫二号应用任务地面数据处理与服务系统的控制单元, 包括服务端容器、工作流管理引擎、高并发工作流调度器等几个部分^[10,11]。系统运行管理与控制通过扩展工作流引擎与工作流调度算法, 集成了包括标准数据产品生产、定制与高级数据产品生产、数据产品质量分析与评价等多类共50余个算法模块。

服务端容器基于Apache+Tomcat搭建, 实现服务端数据处理、存档管理、可视化等各类接口的即时发布, 完成从集群端并行处理程序到标准WebService接口的高效转换。

工作流管理引擎通过算法注册、模块注册、抽象工作流定义、资源映射等方式, 实现多载荷数据处理流程的定义与编排, 在工作流定义过程中根据业务流程设计工作流逻辑, 优化流程中数据文件与参数信息输入, 减少不必要的重复拷贝; 并通过系统的Client端与Server端数据库进行交互, 实现用户对算法参数的按需定制。

高并发工作流调度器确保系统多载荷数据产品生产工作流的高效调度, 通过关系型数据库对各种载荷数据处理算法的基础模块、多个模块的组合所构成的流程进行编排、管理, 在生产不同类型、不同级别载荷数据产品时调用相应的算法流程; 通过优化PBS+MAUI的并发流程调度算法, 实现集群系统的多计算节点负载均衡, 避免出现单点负荷过重、任务队列长时间等待等问题, 如图9所示。

图 9 (网络版彩图)工作流并发调度界面

Figure 9 (Color online) The main interface of parallelized workflow scheduling.

5.4 数据质量分析与控制

5.4.1 数据质量分析评价

针对天宫二号宽波段成像仪、三维成像微波高度计和多波段紫外临边成像光谱仪的成像原理和数据特性,建立了天宫二号对地观测数据产品质量评价算法库,包括辐射、几何、光谱及综合共4类20余种评价指标及其算法,如表1所示。针对天宫二号宽波段成像仪和三维成像微波高度计,开展了地面在轨定标实验,配套开发了宽波段成像仪和微波高度计两个定标资料处理软件,通过数据质量评价结果比较,对处理与产品生产的处理参数和定标参数进行优化调整,实现数据产品质量的控制与提升。

5.4.2 数据质量控制

数据质量控制主要通过外场定标实验的手段来实现。研制的宽波段成像仪定标资料处理软件主要用于实现对宽波段成像仪可见近红外、短波红外、热红外3个谱段18个通道的外场辐射定标资料的自动处理,生成各通道辐射定标系数,并提供对不同时期定标资料的存档管理、定标系数筛选分析功能。利用外场定标实验获取的辐射定标系数,经与MODIS同步验证,可见近红外、短波红外波段的辐射偏差均在7%以内,热红外波段精度在2 K以内,有效地控制了数据辐射产品质量。界面如图10所示。

表1 天宫二号对地观测数据产品质量评价指标列表

Table 1 The quality observation indicators list of Tiangong 2 for Earth observation data products

载荷	评价内容	评价指标
宽波段成像仪	辐射质量	信噪比、清晰度、对比度、信息熵、辐射均匀度、绝对辐射精度
	几何精度	波段配准精度、平面定位精度、影像内精度
	综合质量	综合质量评价
三维成像微波高度计	辐射质量	相对辐射精度、绝对辐射精度、辐射分辨率
	几何精度	平面定位精度、高程精度
	综合质量	综合质量评价
多波段紫外临边成像光谱仪	辐射质量	信噪比、信息熵、对比度
	光谱质量	光谱稳定性、光谱准确性
	综合质量	综合质量评价

图10 (网络版彩图)宽波段成像仪定标资料处理软件界面图

Figure 10 (Color online) The interface diagram of the calibration software for the wide-band imaging spectrometer.

研制的微波高度计定标资料处理软件主要完成微波高度计外场定标数据的录入,包括定标场布设设备参数、定标场二维和三维遥感影像,实现对微波高度计的辐射定标资料与几何定标资料处理以及质量评定,获取辐射定标参数以及几何定标参数等。利用获取的定标系数,三维成像微波高度计后向散射系数图像的绝对辐射精度达到2.0 dB的指标要求;定位误差由校正前的37 m达到校正之后的21 m。界面如图11所示。

5.5 空间应用数据共享服务平台

针对海量空间科学数据管理与共享服务需求,在分析空间科学数据特点的基础上,研发构建了海量空

图11 (网络版彩图)微波高度计定标资料处理软件界面图

Figure 11 (Color online) The interface diagram of the calibration software for the Interferometric Imaging Radar Altimeter.

间科学数据在线服务的载人航天空间应用数据共享服务平台(<http://www.msadc.cn>, 以下简称共享服务平台). 该平台整合了天宫一号、天宫二号、天舟一号等多任务多载荷空间科学数据产品资源, 提供多种形式的数据库检索、下载服务, 满足各类用户对空间科学数据的多样化需求. 共享服务平台门户网站如图12所示.

共享服务平台采用了基于内存查找树的文件快速检索、基于活跃度的动态存储调整、基于临时存储区的可回滚数据归档等数据管理服务策略以及freemarker模板引擎、内容检索等web技术, 以元信息、树形结构、地图、FTP等多种数据检索、下载方式满足科研、公众等多用户群体对多任务、多载荷空间科学数据的复杂数据需求. 平台采用线上与线下相结合的身份认证机制, 控制数据分发权限, 保证空间科学数据使用质量, 促进成果产出.

目前共享服务平台已纳入国家基础科学数据共享

图 12 (网络版彩图)空间应用数据共享服务平台

Figure 12 (Color online) The space application data promoting service platform.

服务平台, 可支持在线并发访问量 ≥ 1000 人, 并发下载线程 ≥ 100 个. 在2016年9月~2019年7月期间, 平台访问量 ≥ 800 万次, 已为200多家高校、科研院所和个人用户提供超过48 TB数据产品, 为数据产品的分发共享和成果产出提供了重要的技术支撑.

6 系统关键技术与解决方案

6.1 基于虚拟相机模型的宽波段图像拼接技术

目前常用的图像拼接算法主要思路是利用图像像素信息(灰度信息、特征点等)进行图像拼接, 该类算法存在不稳定且速度慢的问题, 无法满足快速、高质量的宽波段图像处理要求. 为了在不降低空间分辨率的条件下, 提高影像幅宽, 稳定、高效地对宽波段图像进行拼接处理, 我们创新性地提出了一种基于大幅宽虚拟相机模型的影像自动拼接方法, 主要流程:

- (1) 对多个单相机的严密几何成像模型进行推导、拟合成一个虚拟大幅宽相机的成像模型;
- (2) 利用单相机的严密几何成像模型将影像像点 (s, l) 投影得到地面点坐标 (x, y, z) ;
- (3) 将地面点坐标利用虚拟相机的成像模型投影到虚拟相机焦平面坐标下得到对应像点坐标 (s', l') ;
- (4) 建立从像点 (s, l) 到像点 (s', l') 的坐标映射关系, 将各单相机影像投影到虚拟相机焦平面上, 获得各单相机影像在虚拟相机焦平面坐标下的拼接图像.

影像拼接精度在0.3个像元内, 如图13和14中方框所示. 该算法不仅有效解决了常规图像拼接算法存在的不足, 而且没破坏影像间的几何位置关系, 在几何校正中不会带来额外的定位误差.

6.2 基于内陆水体区域的干涉系统参数定标技术

干涉定标获取载荷在轨工作时的精确干涉参数及

图 13 (网络版彩图)宽波段可见近红外3个相机原始图像

Figure 13 (Color online) The original wide-band imaging spectro-meter image.

图 14 (网络版彩图)宽波段可见近红外3个相机图像拼接结果

Figure 14 (Color online) The mosaic result of the wide-band imaging spectrometer image.

其变化,是三维成像微波高度计数据处理的一项关键技术^[12]。现有的干涉雷达高度计干涉定标方法研究主要集中在海洋定标上,而内陆水体区域的干涉定标无法直接使用海洋交叉定标方法,只能由海洋定标结果插值延拓获取,无法满足实际需求。

针对上述问题,通过研究微波高度计的干涉相位与干涉系统参数之间的关系,我们创新性地提出了一种基于内陆水体的高精度干涉基线定标方法。该方法利用内陆大面积湖泊的单次观测数据和地面实测水位数据获取干涉相位误差,并根据干涉相位误差随视角的变化规律实现干涉基线变化估计。该方法与海洋交叉定标方法相比,具有不受定标区域纬度位置影响、定标性能稳定的优势,利用位于北京附近的官厅水库进行定标实验^[13],经过干涉基线变化校正后的湖泊水位在1 km空间分辨率上的相对反演精度达到40 cm,有效补偿了基线变化引入的系统误差。

6.3 紫外临边成像光谱仪切点高度定位技术

紫外临边成像光谱仪成像方式新颖,大气临边切点高程定位是数据处理中的关键技术。针对此问题,我们创新性地提出一种基于虚拟椭球体的大气临边切点高程定位算法^[14],综合利用平台轨道、姿态和像素视场分布等几何成像参数,在虚拟椭球体上建立了观测视场与地球临边切点的几何模型,实现大气临边切点高度定位,保障了紫外临边观测臭氧等大气廓线的反演精度。

由于无法直接在大气临边高度处选取控制点进行定位误差分析,采用间接法进行验证。选取目标点经纬度为(20.982, -42.159)处,随高度变化反演的O₃产品廓线,与美国Suomi NPP卫星搭载的臭氧观测设备(OMPS)的臭氧反演曲线交叉验证,临边切点高度定位

误差在1个像素内,如图15所示。

6.4 业务流程高并发调度与管控技术

由于多载荷数据产品生产算法的并行化方法不同,使用单一的CPU或GPU计算资源无法满足数据产品的快速生产需求;同时,传统的数据密集型 workflow 调度算法可根据计算任务优先级、先进先出等策略进行作业调度,实现批量任务的并行调度,但针对混合计算资源的调度效率不高,无法高效利用CPU与GPU的混合计算能力。

在多源海量数据并行处理方面,提出了PBS+MAUI高并发流程调度算法,基于CPU、GPU构建混合资源池,分别研发基于MPI与CUDA的并行化模型,实现了宽波段数据、紫外临边数据几何校正的CPU+GPU混合计算资源的高效调度与管控。

对宽波段成像仪数据采用MPI并行处理,以标准景作为最小处理单元的同时,利用计算节点间消息传输接口同步重采样处理标准景内分块数据,提高标准景数据产品的并行处理效率;对紫外临边数据以标准帧为最小单元采用GPU并行处理,解决了紫外临边数据单轨成像时间长、数据量大的问题,最大化发挥GPU在多核处理上数据高并发处理的计算优势,日数据处理能力大于1.5 TB,混合计算能力比仅使用CPU计算资源时提高4倍以上,如图16所示。

图 15 (网络版彩图)紫外临边与OMPS反演O₃产品的定位误差对比

Figure 15 (Color online) Comparison of positioning errors between multi-azimuth UV imager and OMPS inversion O₃ products.

图 16 (网络版彩图) 基于CPU与GPU的紫外环形数据并行处理加速比

Figure 16 (Color online) Speed-up ratio of the ultraviolet ring imager data parallelized processing based on CPU and GPU.

7 总结与展望

天宫二号应用任务地面数据处理与服务系统已稳定高效运行近三年, 是一套集高精度数据并行处理、高效数据归档管理、数据产品可视化、数据质量分析与控制和数据分发共享等于一体的高性能地面数据系统, 系统性地解决了多载荷机制新、技术先进、实验机理差异大引起的数据处理算法复杂多样问题; 解决了应用研究领域差异大, 研究人员对不同类别/级别数据产品需求多样、精度要求高、处理时效性快等问题; 解决了长期业务运行的数据产品质量分析与评价问题; 攻克了数据类型多、体量大造成的数据存档管

理难度大问题; 克服了数据在线服务的便捷性、时效性低等相关技术问题。该系统累计处理、存储、发布科学实验和应用试验数据量超过200 TB, 为国内外广大科研工作者累计提供了超过48 TB的数据产品, 这些数据及对应数据产品在空间材料、空间天文与空间物理、空间基础物理、空间地球科学及应用的国土资源、海洋、湖泊、生态环境、农业等众多领域开展了应用研究, 所生产的数据产品和提供的数据服务满足各领域研究与应用的需求, 为我国天宫二号空间科学与应用数据的应用与推广发挥了重要的系统支撑作用。

目前该系统基于高性能计算系统实现了多领域、多载荷、海量数据的集中处理、管理和由数据中心单向的提供数据产品的分发服务, 如何结合云计算平台, 利用容器技术实现资源的并行调度与利用, 如何开发智能化的技术方法实现数据更精准、更大范围的共享共用, 还有待进一步提升。未来我国载人空间站阶段, 地面数据系统将会在现有技术基础上, 深入结合云计算、大数据和人工智能等技术, 构建地面数据智能化云服务系统, 提供弹性可伸缩的计算、存储资源与应用研究运行环境, 形成数据存、算一体化、“从数据服务到成果共享”的循环生态模式, 提高数据处理、管理、服务的应用效率, 提供可定制化精准服务, 为高效数据运营、大数据智能服务、数据挖掘与分析等业务提供技术保障, 为空间站任务科学成果的产出提供更高水平的系统技术支持。

参考文献

- 1 Lynnes C S, Baynes K E, Mcinerney M. Earth Observing System Data and Information System (EOSDIS). In: Van Nostrand's Scientific Encyclopedia. New York: John Wiley & Sons, Inc., 2005
- 2 Osuna P. ESA science archives at the European space astronomy center. In: Joint European & National Astronomy Meeting, JENAM 2010. 2010
- 3 Qian J M, Sun A L, Xu Z, et al. Fengyun series meteorological satellite data archiving and service system (in Chinese). J Appl Meteorol Sci, 2012, 23: 369–376 [钱建梅, 孙安来, 徐喆, 等. 风云气象卫星数据存档与服务系统. 应用气象学报, 2012, 23: 369–376]
- 4 Buyya R. High Performance Cluster Computing: Architectures and Systems, Volume 1. Upper Saddle River: Prentice Hall, 1999
- 5 Zhang W, Wang L, Ma Y, et al. Design and implementation of task scheduling strategies for massive remote sensing data processing across multiple data centers. Softw Pract Exper, 2014, 44: 873–886
- 6 Liu Z, Liu Y, Wang B. High-precision geometric correction of Tiangong-2 interferometric imaging radar altimeter. In: Proceedings of the Tiangong-2 Remote Sensing Application Conference. Beijing, 2018. 104–111
- 7 Kong W, Chong J, Tan H. Performance analysis of ocean surface topography altimetry by Ku-Band Near-Nadir interferometric SAR. Remote Sens, 2017, 9: 933
- 8 Tan H, Li S, Liu Z, et al. Evaluation of several filtering and unwrapping methods for the interferometric imaging radar altimeter. In: Proceedings of the Tiangong-2 Remote Sensing Application Conference. Beijing, 2018. 94–103

- 9 Gu Y, Wang Q, Jia X, et al. A novel MKL model of integrating LiDAR data and MSI for urban area classification. *IEEE Trans Geosci Remote Sens*, 2015, 53: 5312–5326
- 10 Mashayekhy L, Nejad M M, Grosu D, et al. Energy-aware scheduling of MapReduce jobs for big data applications. *IEEE Trans Parallel Distrib Syst*, 2015, 26: 2720–2733
- 11 Bikas S, Hitesh S, Siddharth S, et al. Apache Tez: A unifying framework for modeling and building data processing applications. In: *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*. Melbourne, 2015. 1357–1369
- 12 Lu H L, Li P, Li Y W, et al. Error correction and calibration of the off-nadir direction of interferometric altimeter (in Chinese). *Guid Fuze*, 2016, 37: 28–33 [卢护林, 李鹏, 李银伟, 等. 干涉高度计侧视方向误差修正与定标方法. *制导与引信*, 2016, 37: 28–33]
- 13 Tan H, Li S Y, Liu Z W. An assessment of spaceborne near-nadir interferometric SAR performance over inland waters with real data. In: *ISPRS TC III Symposium on Developments, Technologies and Applications in Remote Sensing*. Beijing, 2018. 1635–1638
- 14 Liu Z W, Li S Y. High precision limb height calculation of ultraviolet forward spectrometer of TianGong-2 (in Chinese). *Opt Precis Eng*, 2018, 26: 1517–1523 [刘志文, 李盛阳. 天宫二号紫外前向光谱仪大气临边高度计算. *光学精密工程*, 2018, 26: 1517–1523]

Tiangong-2 application mission ground data processing and service system

LI ShengYang^{1,2}, LIU ZhiWen^{1,2}, ZHANG WanFeng^{1,2}, TAN Hong^{1,2}, LIU YunFei^{1,2}, WANG Bo^{1,2} & GU YiDong¹

¹ *Technology and Engineering Center for space Utilization, Chinese Academy of Sciences, Beijing 100094, China;*

² *Key Laboratory of Space Utilization, Chinese Academy of Sciences, Beijing 100094, China*

Driven by the demand for effective data processing, archive management, quality analysis, and operational services, as well as the requirement that data applications utilize the payloads of Tiangong-2 with various mechanisms that are designed for advanced scientific experiments and earth observation, an integrated and ground-based data processing and service system with comprehensive support capabilities and intuitive technology features was developed. The system enables multisource and automated massive data-parallel processing. It has high-precision and high-quality production and enables efficient data archiving. It also facilitates visual management, data quality analysis, and control, online distribution services, and operation, etc. This paper (1) summarizes the current situation and development trend of ground processing and service systems for space data; (2) analyzes the technical requirements in processing, management, and sharing of space data for Tiangong-2 applications; (3) systematically introduces the overall architecture, operation processes, functions, and interfaces of the ground data system; (4) designs and implements in detail the functions of data-parallel processing and product production, visual-data storage and management, system scheduling and control, data quality analysis and control, and data online sharing services; (5) further analyzes key technical problems of high-precision data processing and concurrent dispatching of business processes for new earth observation payloads; and (6) elaborates the specific solutions. The system has been running steadily for 3 years and has provided more than 48 TB data products for researchers around the world. The ground data-processing system has played an important role in supporting scientific researches, outputs, and application promotion in various application fields of China Manned Space Engineering.

Tiangong-2, ground data system, data processing and service, data distribution and sharing

doi: [10.1360/SST-2019-0331](https://doi.org/10.1360/SST-2019-0331)