国家高性能计算环境的虚拟数据空间运行支撑技术研究

何小雨[1,2]，邓笋根[1]，栾海晶[1,2]，牛北方[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 100190

摘要
国家高性能计算环境的特点是广域分散、系统异构、存储资源隔离自治，这对存储设备管理、数据跨域共享等提出了极大的挑战。首先阐述了虚拟数据空间系统的概念，然后分析了其作为国家高性能计算环境的一部分，如何通过视图访问、数据共享、计算环境对接有效降低跨域的访问开销；接着通过模块化方式将虚拟数据空间系统与国家高性能计算环境进行深度的融合，将其功能补充到国家高性能计算环境中；最后通过统一的虚实空间用户管理框架实现了跨域统一、透明安全的存储服务和对大型计算应用的支撑，这对于国家高性能计算环境的发展具有十分重要的意义。

关键词
虚拟数据空间系统；国家高性能计算环境；数据分布；用户视图；环境集成

中图分类号：TP31 文献标识码：A doi: 10.11959/j.issn.2096-0271.2021019

Study of technique support on the operation of virtual data space in national high performance computing environment

HE Xiaoyu[1,2], DENG Sungen[1], LUAN Haijing[1,2], NIU Beifang[1,2]
1. Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China
2. University of Chinese Academy of Sciences, Beijing 100190, China

Abstract
The national high performance computing environment (NHPCE) is characterized by wide area dispersion, heterogeneous system and storage isolation autonomy, which posed great challenges to storage management and data sharing. Firstly, the concept of virtual data space system was described. Secondly, how the virtual data space system effectively reduce the cost of cross domain access through viewing-access, data sharing, computing environment docking , was analyzed. Thirdly, the virtual data space system were deeply integrated with the NHPCE by means of modularization and its functions were added to the NHPCE. Finally, the unified virtual and real space user management framework was used to achieve cross domain unified, transparent and secure storage service and the support of large-scale computing application, which is of great significance to the development of NHPCE.

Key words
virtual data space system, national high performance computing environment, data distribution, user view, environment integration

1 引言

国家高性能计算环境(nation high performance computing environment, NHPCE)广域分散、系统异构的特点显著增加了虚拟数据空间存储设备的管理难度[1]。针对该问题,国家高性能计算环境中虚拟数据空间运行支撑技术课题组拟研究国家高性能环境节点部署技术。通过对高性能计算环境典型文件系统的分析,采用针对性挂载方式,在不修改本地系统的情况下,实现虚拟数据空间中跨域存储资源的高效接入,从而有效地降低虚实空间界面间的开销。

虚拟数据空间系统的目标是集成两个国家网格主节点(中国科学院计算机网络信息中心、上海超级计算中心)和3个国家超级计算中心(广州、济南、长沙)的数据存储资源,每个中心需提供不少于300 TB的共享存储空间。虚拟数据空间系统部署成功后,用户提供可在统一的虚拟数据视图上运行上层的应用。用户负责选择具体应用,发送应用数据到统一数据空间,然后提交应用任务。在虚拟数据空间软件部署运行和应用验证优化方面,项目团队涵盖上述3个国家级超级计算中心以及两个国家网格主节点,拥有“天河”“神威”“元”等超级计算机系统,总计算能力超过100 PFlops,在线存储量近30 PB,活跃用户超过6 000个,支撑数值模拟、大数据处理、人工智能等众多大型计算应用,为项目开展虚拟数据空间的软件部署运行和应用验证优化工作提供了资源和应用基础。

国家高性能计算环境的虚拟数据空间运行支撑技术研究包括:虚拟数据空间与国家高性能计算环境接口技术、计算与存储协同调度技术、虚拟数据空间软件部署及验证方法等。研究虚拟数据空间与国家高性能计算环境接口技术,实现虚拟数据空间与国家高性能计算环境的对接;突破计算与存储协同调度技术,实现数据与计算任务的全局优化调度;研究虚拟数据空间软件部署及验证方法,提高软件部署效率,从而实现虚拟数据空间构建、数据共享迁移等技术的集成,构建一个可运行于国家高性能计算环境的虚拟数据空间系统,并可被外部应用访问。

2 国内外现状

国家高性能计算环境又称中国国家网格[2-6],其发展建设最早可追溯至20世纪90年代末。21世纪初,国家863计划出资建设了5个高性能计算中心,所构建的高性能计算环境是未来国家网格环境的初级形态。随后,该国家网格环境在我国“五年计划”的大力支持下得到了长足的发展。目前,中国国家网格已涵盖了众多国内外优质的高性能计算资源,可为普通用户提供高效便捷的计算服务,同时为基于多领域计算平台的应用建设提供了有力的支持[7-8]。随着中国国家网格的发展进步,通用计算平台、应用社区等应运而生[9]。近年来,在大数据技术的推动下,大规模数值计算的应用不断深入大型异构并行系统,这种需求也反过来推动了高性能计算环境的发展,解决了计算与存储的协同问题,使得更多富有挑战性的任务的解决成为可能。除此之外,为了满足快速增长的用户计算需求,中国国家网格的底层关键模块在不断地进行功能扩展和优化升级,以满足在不同应用领域中用户对环境扩展性、易用性和可靠性的需求。

1983年,美国国家科学基金会、国防

部和能源部等多个部门和组织曾联合向政府提出大力发展科学和工程计算工作的报告。此外，美国政府高度重视高性能计算项目的发展，并于1991年提出了“高性能计算与通信计划”。该项目的目标是通过深入研究高性能计算环境解决一系列科学难题。1994年，美国能源部开展了为期10年的“加速战略计算创新”计划，该计划通过运用超级计算机极大地加速了其核武库的建设发展。1998年，美国能源部倡议在全国范围内实施“科学模拟”计划，提出要加速“燃烧系统”与“全球气候系统”的科学模拟研究。2002年，美国国防部启动了“高生产率计算系统”计划，规划了未来20年内超级计算机体系结构的发展，并计划分阶段有序实施。

美国极限科学与工程发现环境（extreme science and engineering discovery environment，XSEDE）的用户门户提供了多样化的计算和技术支持服务，借助Globus[10]实现大规模的数据传输。XSEDE的目标是在TeraGrid[11]的基础上建立一个可以提供私密安全环境的网络基础设施，使研究人员可以获得所需资源、服务和合作的支持。例如，印第安纳大学负责开发科学网关、在线工具和门户，使科学家更易于访问先进计算资源。此外，印第安纳大学还负责提供虚拟机、网络监控和备份操作等服务。田纳西大学辅助创建美国高性能计算机与研究设施之间的新一代链接。新的XSEDE超级计算机网格将创造出强大的工具来解决部分高度复杂的科学问题，如通过气候建模、药物开发、DNA排序和各类模拟来解决气候变化、不治之症和能源危机等问题。此外，由田纳西大学和橡树岭国家实验室联合运作的国家计算科学研究所负责改善高性能计算机、数据资源和实验设施之间的链接。

欧洲网格基础设施（European grid infrastructure，EGI）是一个可持续的泛欧基础设施。EGI的愿景是让所有学科的研究人员通过简单、完整和开放的渠道获取先进的数字功能、资源和专业知识，以及进行计算和数据密集型科学研究和创新。EGI整合了各类资源，主要包括各研究机构及国家的数字能力、数字资源和专业知识，从而为研究基础设施创造并提供解决方案。截至2016年9月，EGI提供了826 500颗CPU核用于高通量计算、6 600颗CPU核用于云计算，在线存储容量达到285 PB，档案存储容量达到280 PB。EGI的前身是欧洲数据网格（European data grid，EDG）和欧洲科研信息化网格（Enabling grid for e-science in Eruopean，EGEE），它们实现了对计算、存储和网络资源的跨国访问。然而，原有的整套服务是根据早期科学团体的需求量身定制的，并不总是能满足新团体的需求，因此启动了EGI。EGI于2010年正式启动，第一阶段为面向欧洲科研人员的集成可持续泛欧基础设施（EGI-InSPIRE）项目，旨在创建一个无缝系统，满足当前和未来的科学工作需求。2014年12月EGI-InSPIRE项目结束。2015年3月进入第二阶段，为促进EGI社区迈向开放科学公地，EGI-Engage项目启动，旨在扩展欧洲在计算、存储、数据、通信、知识和技能方面的重要联合服务能力，加速开放科学共享的实施，以补充特定社区功能。

欧洲WLGG（worldwide LHC computing grid）是世界范围内最大的高能物理计算和存储设施，通过专门的文件服务，使用现有的协议（如HTTP、FTP和GridFTP等）提供文件传输和共享服务。WLCG创建于2002年，是一个全球性计算机中心合作项目，旨在提供资源，用于存储、分发和分析大型强子对撞机（large hadron collider，LHC）每年产生

的几十PB的数据。2016年，大型强子对撞机产生的数据在过滤99%以后，还能达到50 PB，该数据量相当于1 500万部高清电影的总数据量。WLCG由欧洲核子研究组织（European Organization for Nuclear Research）负责协调，连接着全球42个国家的170多个计算中心，以及数个国家与国际网格，每天可运行200万个作业，是当今世界上规模较大的网格计算环境之一。通过部署一个覆盖全球范围的计算网格服务，WLCG项目将欧洲、美洲、亚洲等地区的超级计算中心集成到一个虚拟的计算组织中，为大型强子对撞机实验提供计算资源，包括CPU计算资源、数据存储能力、处理能力、传感器、可视化工具、网络通信设施及其他资源等。实验产生的数据分布于全球，欧洲核子研究组织会对原始数据进行备份。数据经过原始处理后，将在计算网格全天候运行的支持下分布式地存储到欧洲、北美和亚洲的13个高水平研究中心，再从那里分散到世界各地上百个研究中心，由全世界多位物理学家合作处理实验数据。

3 国家高性能计算环境接口研究

高性能计算（high performance computing，HPC）任务一般具有以下特征：

- 大规模并行，可上升至百万核；
- 长时间计算；
- 稳定的网络通信和较大的存储；
- 计算任务以批处理作业的方式完成，且从任务开始运行直到结束，用户不可交互。

高性能计算环境提供给用户的使用方式主要包括图形用户界面（graphical user interface，GUI）、命令行（command line）和网站门户（Web portal）3种。其中，GUI在界面友好性、功能完善性和响应速度等方面革新了人机交互的方式，但需要在用户的计算机上安装软件。命令行方式操作灵活、功能全面，但需要用户具备较高水平的编程能力。基于Web portal的HPC使用方式结合了前两种方式的优点，用户可在Web图形界面上完成作业提交和管理，由于其简单易用性，这种使用方式已经成为当前HPC使用方式的潮流，如中国科学院超级计算环境、中国国家高性能计算环境、美国TeraGrid Portal群及其后续项目等都使用这种方式。

国家高性能计算环境普遍采用基于表述性状态转移（representational state transfer，REST）[11-13]风格的应用程序接口（application programming interface，API）。该API是连接环境系统软件和计算应用服务平台的中间层，负责将各项请求信息传递到异构计算资源，并且通过执行操作来查询和获取信息，包括虚拟数据空间中计算数据（输入/输出数据）的分布信息（存储信息）。API负责以合适的形式封装环境系统软件提供的功能，并优化已有API的性能和效率。

在接口设计中，重点设计高性能计算任务中虚拟数据空间系统提供的多副本数据服务，包括大规模数据的快速传输、文件管理等功能接口；然后研究基于虚拟数据空间的多数据来源的作业提交、共享输入数据和共享应用软件的批量作业管理等；在此基础上，进一步研究高性能计算环境的工作流任务模型，包括作业提交、状态管理、数据传输、输入文件格式转换、输出文件质量检查和依赖条件判断等，并设计可扩展的插件机制，利用已有成果完成作业调度、前后处理等功能，从而为不同学科领域的应用计算和上层应用社区提供定制化的开发功能。

高性能计算环境接口包括：机群节点

列表查询接口、应用列表查询接口、环境队列查询接口以及作业系统相关接口等。

高性能计算环境接口一方面在高性能计算环境中为虚拟数据空间提供统一的全局虚拟数据空间的数据访问服务，另一方面为高性能计算环境中数据与计算的协同调度提供相应的服务。除此之外，虚拟数据空间集成部署后，将在高性能计算环境中使用，因此还设计了相应的接口供高性能计算环境调用。高性能计算环境接口和虚拟数据空间接口分别见**表1**和**表2**.

在接口设计过程中，对JSON响应输出格式进行了内容规范，具体如下：

- 如果交互正常，则返回的HTTP status code为200，其他错误内容参见HTTP的标准定义，以后将有详细的编码定义；
- 调用HTTP API时，平台返回的响应输出由status_code、status_msg两个参数组成，分别用于描述错误码和错误信息。status_code为0，表示操作成功执行，其他错误码参见具体函数的说明；
- 交互正常（http_code=200）时，响应会生成request_id字段，该字段的值由服务端生成，并返回给用户，以便进行问题追查与定位。

研究中使用response_params表示API返回的结果，其是由n个包含<key,value>对的元素组成的JSON对象。

表1 高性能计算环境接口

接口名称	接口描述	接口类型	请求URL	请求类别	请求说明	参数
vhpcs	查询高性能计算环境提供的HPC列表信息	REST	/resources/hpcs	GET	HPC信息查询	无
vapps	查询高性能计算环境提供的应用信息	REST	/resources/applications	GET	计算环境提供的应用信息查询	host：可选；字符串；表示计算环境中HPC的名称
vqueues	查询高性能计算环境中的应用可以使用的计算队列信息	REST	/resources/applications/{appName}	GET	appName表示应用的名字	host：可选；字符串；表示计算环境中HPC的名称
vjobs	查询高性能计算环境中的作业信息，包括所有已经提交的作业	REST	/jobs	GET	作业信息查询	host：字符串；表示计算环境中HPC的名称，被查询的作业在该HPC上计算任务

表2 虚拟数据空间接口

接口名称	接口描述	接口类型	请求URL	请求类别	请求说明	参数
vlu	查询虚拟数据空间用户信息	REST	/users/search/:name	GET	账户信息查询	账户名
vlg	虚拟数据空间用户组以区域为粒度，即一个区域内的主人和成员为一个用户组，区域主人具有增加/删除成员、修改/查询组的权限的功能	REST	/zone/share	POST	区域共享	区域名、账户名
		REST	/zone/sharecancel	POST	区域共享取消	区域名、账户名
		REST	/auth/modify	POST	区域权限修改	账户ID
		REST	/auth/search	POST	区域权限查询	账户ID
vmkuser	创建虚拟数据空间用户	REST	/users/registration	POST	账户注册	账户注册信息
vmoduser	修改虚拟数据空间用户信息	REST	/users/modify	POST	账户信息修改	账户名
vrsync	同步本地的数据与虚拟空间的数据，由于虚拟数据空间支持POSIX接口操作，故直接调用linux sync命令即可	–	–	–	同linux sync命令	同linux sync命令

4 数据与计算协同调度方法

计算任务的运行时间高度依赖资源的布局和可用性，资源失配将显著增加运行时间。针对该问题，研究数据与计算任务的协同调度机制，基于设计的环境接口、特征模型，动态监视资源状态，对全局资源进行统一管理和分配，设计优化解决方案，以实现数据与计算任务的协同调度。

在高性能计算环境中，各个超级计算中心集群的计算能力、调度系统、部署的应用软件以及网络带宽都有区别，为了避免传统单一的计算任务调度，采用计算与数据协同调度算法，依据计算特征和数据布局选择任务和数据节点，实现在广域范围内高效合理的计算任务分布和数据布局，以降低应用的跨域访问开销。

在传统集群计算机系统中，计算系统与存储系统是分开的，且其资源管理和任务调度系统相互独立[11]。本部分进行了计算与数据协同调度的研究，从**图1**可以看出，计算与数据协同调度算法不仅考虑了高性能计算环境计算服务中的因素，还将结合虚拟数据空间中的与一些数据服务任务相关的数据属性。在高性能计算环境中，系统接收到计算任务的请求时，通过计算与数据协同调度算法选择高性能计算环境的目标集群，同时计算该环境中的计算服务，通过虚拟数据空间系统的数据服

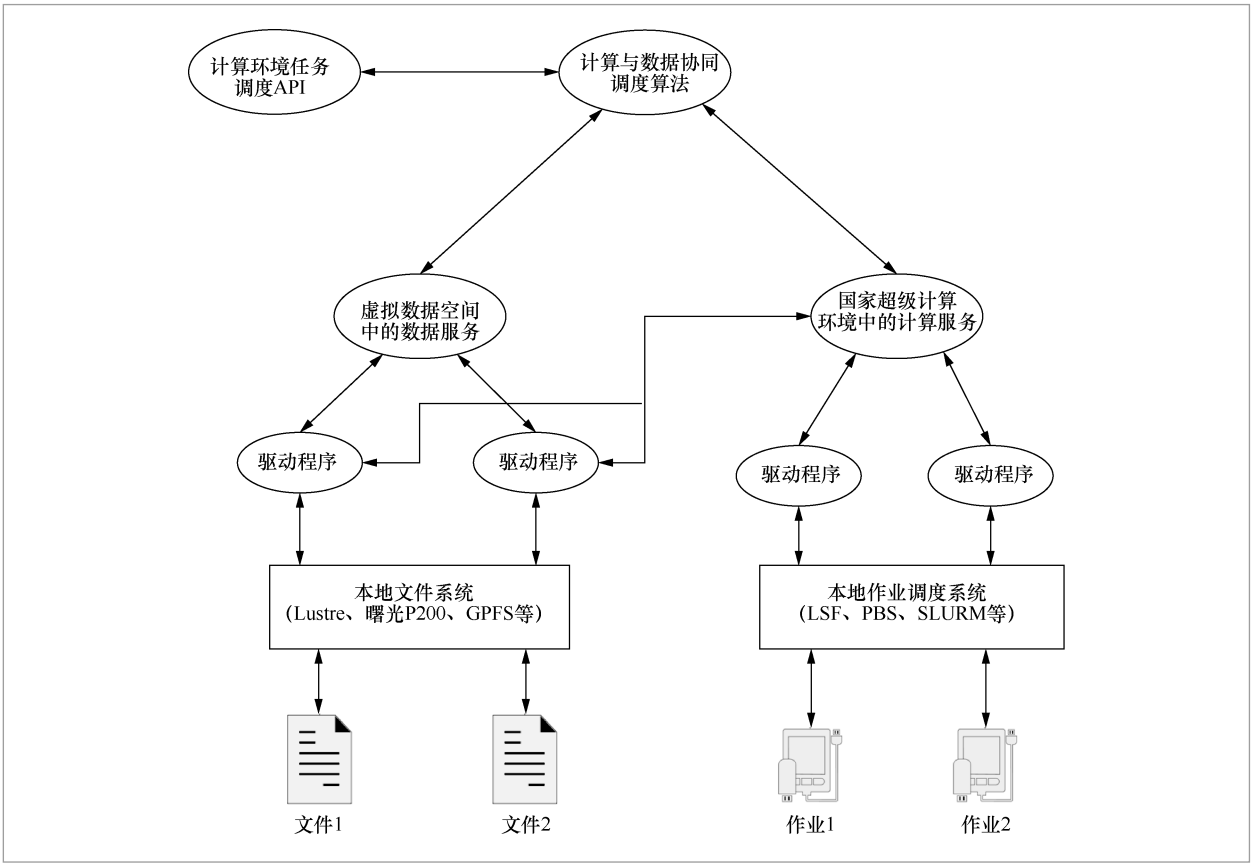

图1 高性能计算环境计算与数据的协同调度系统的结构示意图

务进行计算任务输入/输出的广域的虚拟数据空间中的数据访问。如某个计算任务的执行程序被部署在A、B、C 3个超级计算中心，在分配计算资源时，A、B、C都可以作为该计算任务的目标执行集群，这时需要关注该计算任务的输入数据在分布式虚拟数据空间中的分布情况：首先查看该输入文件是否在A、B、C中存在副本，若有，则在协同调度模型中该项因子的权重生效；判断是否需要在A、B、C中创建该文件的副本，其中需考虑创建各个副本的代价情况；还需要考虑其他因素，如队列、用户权限等。

高性能计算环境计算与数据的协同调度模型充分考虑了高性能计算环境中的计算集群资源及队列情况、数据分布状态，各个超级计算中心之间的网络互连情况、计算任务的需求，以及各个超级计算中心上的应用部署情况。

高性能计算环境计算与数据的协同调度模型如**图2**所示。其首先将接收到的由计算集群、输入输出数据、网络、任务以及应用部署5大类16个小项数据输入参数组成的矩阵输入协同调度模型中，通过参数匹配和参数优化进行任务和数据的调度。

从国家高性能计算环境来看，任务的启动包含数据准备时间、任务在目标中心集群上的调度时间。

基于虚拟数据空间的广域数据的访问，进一步实现环境节点管理、环境任务管理以及环境数据管理等，具体如**图3**所示。环境节点管理部分主要指高性能计算环境与虚拟数据空间环境调用接口之间的交互，针对高性能计算环境中的计算、数据、集群3个方面实现集群节点管理、集群队列管理、集群作业管理和集群应用管理；数据与计算协同调度模块包括环境API、与虚拟数据空间对接的数据空间API，此模块的任务是利用启发式协同调度算法进一步实现数据迁移、任务迁移及任务特征匹配。

（1）对计算资源、存储资源、数据准备（包括数据迁移）、任务迁移以及任务预期运行时间等调度要素进行形式化定义，为接下来的调度模型算法设计奠定基础。

- 应用信息的相关信息如**图4**所示，包括版本、部署的节点等。
- 数据信息的相关信息如**图5**所示，包括大小、分布节点等。
- 队列信息的相关信息如**图6**所示，包括状态、允许的核数、时间、排队情况等。

在协同调度算法设计中，本文做出相应的定义，具体如下。

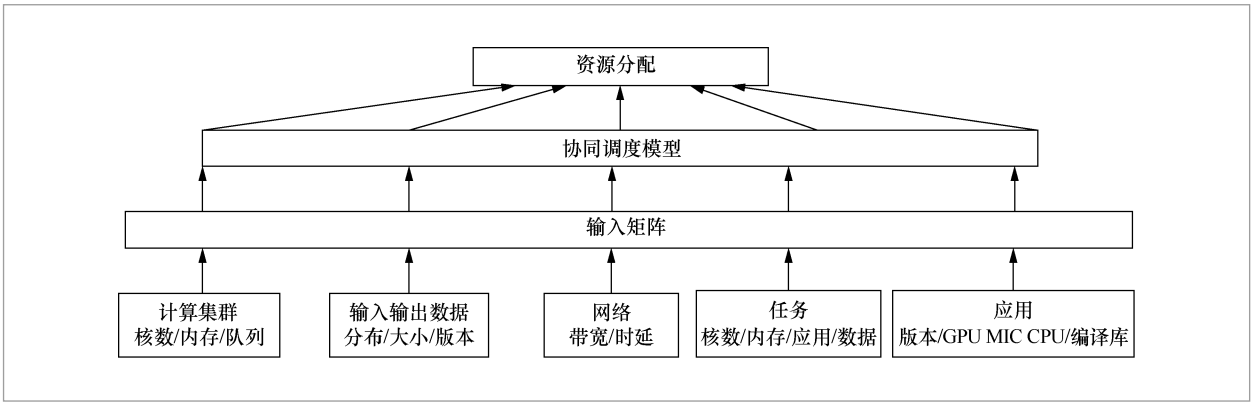

图 2 高性能计算环境计算与数据的协同调度模型

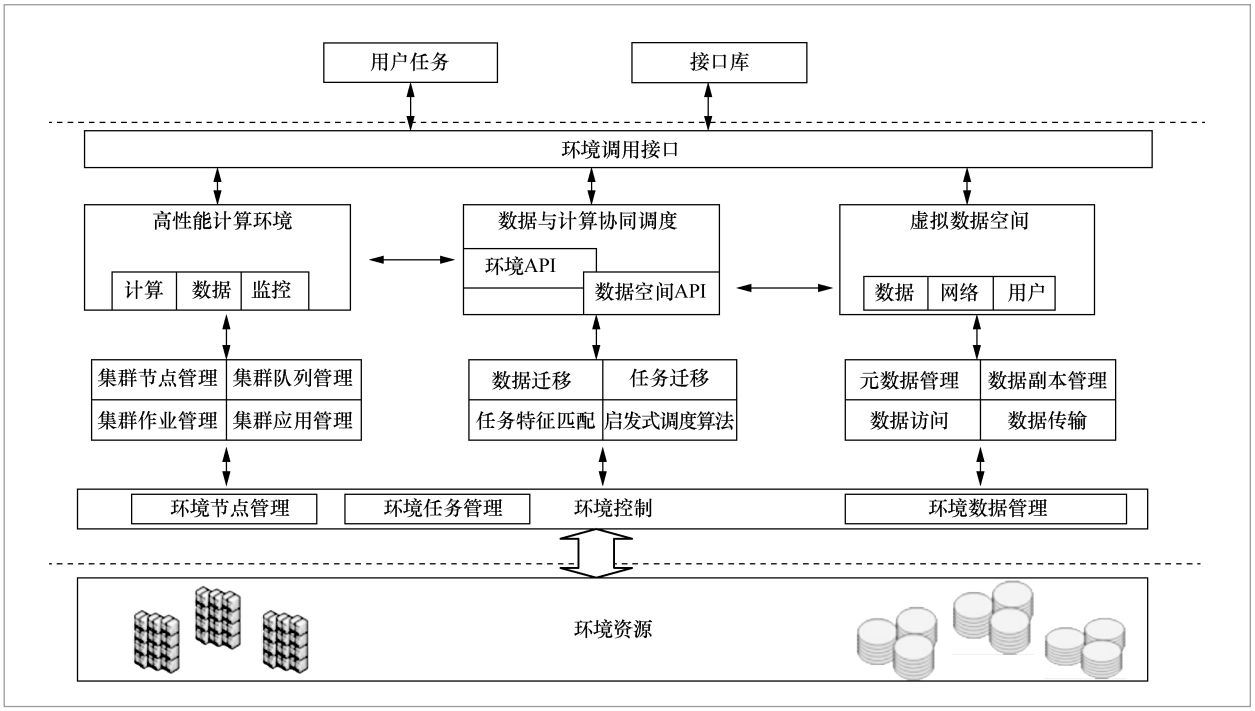

图 3 虚拟数据空间协同调度设计框架

● 应用的定义

$$\mathrm{Node}_{\mathrm{app}}(N_{a1},N_{a2},N_{a3},\cdots) \quad (1)$$

$\mathrm{Node}_{\mathrm{app}}$表示应用在各点的部署情况，其中，$N_{ai}$表示应用迁移情况。

$$N_{ai}=\begin{cases}-1, \text{不可迁移}\\ 0, \text{无须迁移}\\ \text{value}, \text{可迁移}\end{cases} \quad (2)$$

● 数据的定义

$$\mathrm{Node}_{\mathrm{data}}(N_{d1},N_{d2},N_{d3},\cdots) \quad (3)$$

$\mathrm{Node}_{\mathrm{data}}$表示数据分布/迁移的情况，其中，$N_{di}$表示应用迁移代价。

$$N_{di}=\begin{cases}-1, \text{不可迁移}\\ 0, \text{无须迁移}\\ \text{value}, \text{可迁移}\end{cases} \quad (4)$$

● 队列的定义

$$\mathrm{Node}_{\mathrm{queue}}(N_{q1},N_{q2},N_{q3},\cdots) \quad (5)$$

$\mathrm{Node}_{\mathrm{queue}}$表示队列是否满足任务的条件以及任务执行需要的等待时间，其中，N_{qi}表示队列状态。

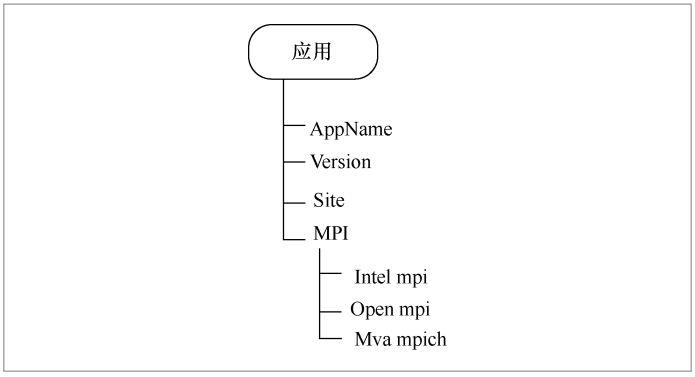

图 4 应用的相关信息

$$N_{qi}=\begin{cases}-1, \text{队列不可用}\\ 0, \text{队列可用无须排队}\\ \text{value}, \text{需要排队}\end{cases} \quad (6)$$

（2）建立协同调度的虚拟任务队列，如**图7**所示。

在全局的虚拟数据空间与高性能计算环境的计算与数据协同调度设计中，采用一种虚拟队列的方式，将环境中各个点的

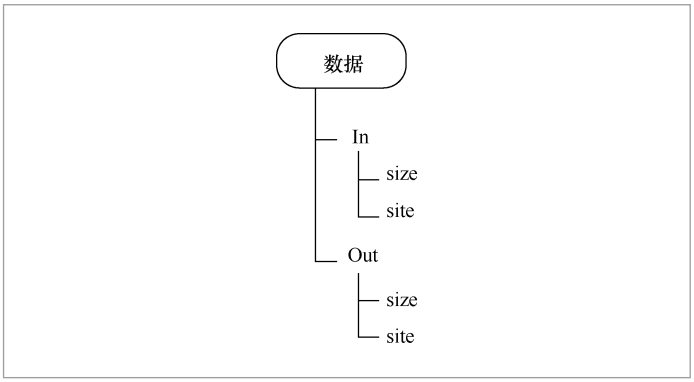

图 5 输入 / 输出数据的相关信息

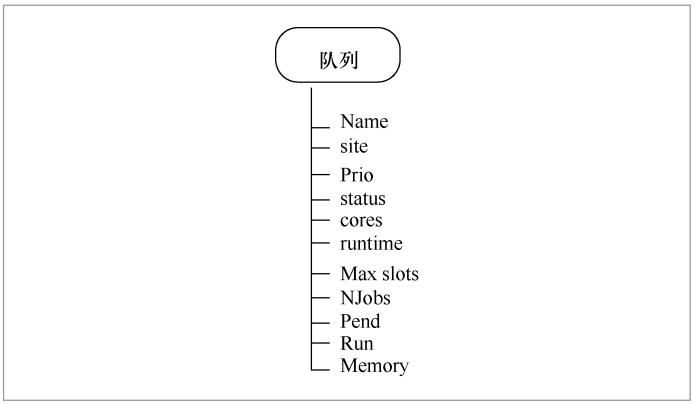

图 6 队列相关信息

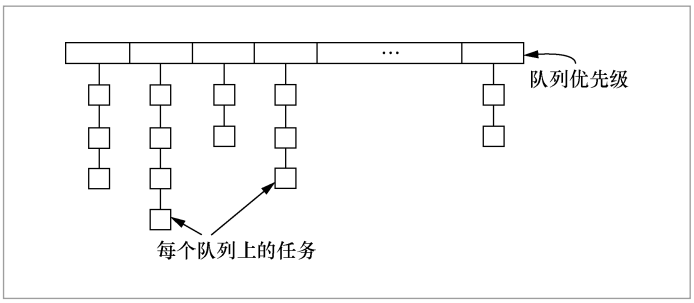

图 7 协同调度的虚拟任务队列

资源虚拟成全局调度的一个队列。根据各个点的资源和任务使用情况设定各个虚拟队列的优先级。

在调度过程中，结合任务的部署情况、数据准备情况、各虚拟队列的任务排队情况，以及任务和数据迁移代价，综合制定环境的调度目标。

5 虚拟数据空间系统与高性能环境的集成

在高性能计算环境的相关接口以及计算与数据的协同调度模块的基础上，将虚拟数据空间接口集成到国家网格环境的服务器中，与高性能计算环境形成完整的系统。

如图8所示，虚拟数据空间系统目前已完成五大超级计算中心的部署，这些中心同时也是国家高性能计算环境的节点，将虚拟数据空间系统服务的客户端节点纳入国家高性能计算环境中，用户在高性能计算环境中使用虚拟数据空间系统提供的功能，实现虚拟数据空间系统对国家高性能计算环境的扩展。

作为国家高性能计算环境的补充，虚拟数据空间系统为国家高性能计算环境提供了广域环境下各中心数据存储资源的统一虚拟访问。本研究将虚拟数据空间系统集成到国家高性能计算环境中的数据与计算服务层，从而为上层的计算服务平台提供广域的虚拟数据空间的数据访问服务。

（1）GVDS接口client命令供CNGrid client命令行使用

通过开发虚拟数据空间系统的接口为国家高性能计算环境提供支持。虚拟数据空间系统与国家高性能计算环境的调用接口，可为上层的计算服务平台中的各类服务提供虚拟数据空间的数据查询/访问/传输。支持从多个输入源选择输入文件或存储输出文件，使得高性能计算环境中的用户能通过虚拟数据空间系统访问广域环境中其他网格节点的数据，实现高性能计算环境中计算任务能够访问跨域分散的存储资源，使用户能够从使用角度看到一个与本地数据空间一致、可提供统一访问与管理的虚拟数据空间，并在国家高性能计算

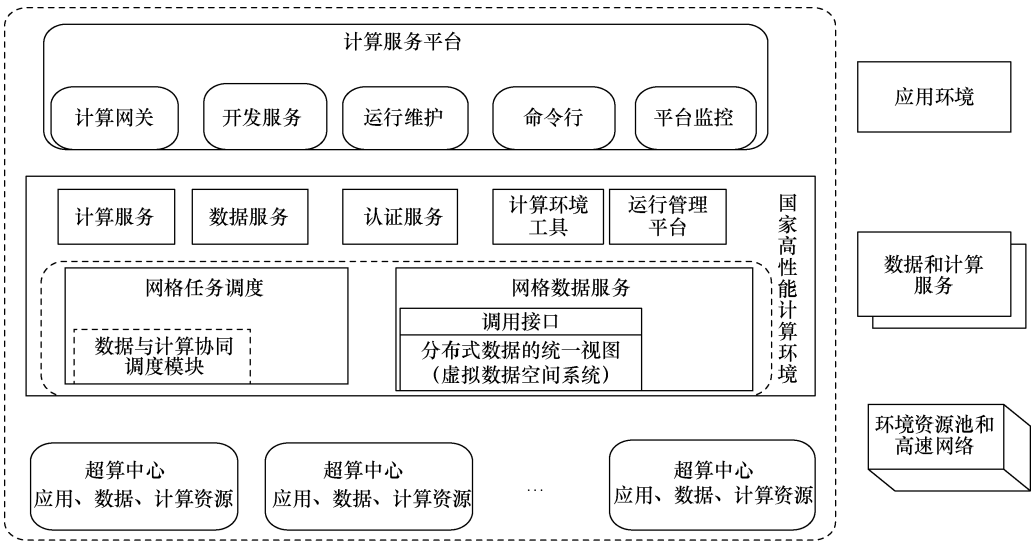

图 8 虚拟数据空间系统与国家高性能计算环境的集成系统结构

环境中为用户的计算任务提供整个环境的计算与数据的协同能力。

将GVDS接口命令部署到高性能计算环境中。用户调用已封装好的命令可实现虚拟数据空间与国家高性能计算环境中的中国科学院超级计算中心、国家超级计算广州中心、上海超级计算中心、国家超级计算长沙中心以及国家超级计算济南中心的数据信息的全局统一访问。

(2) GVDS API供CNGrid portal调用

在天气预报典型应用中，通过命令行方式调用系统中的POSIX接口，实现对虚拟数据空间系统的应用数据的访问。

(3) 虚拟数据空间中用户身份与高性能环境的认证

虚拟数据空间系统作为一个社区，拥有独立的一套用户管理机制，包括账户注册、登录、注销等。CNGrid client命令行与GVDS client命令行都与服务器用户进行一对一映射。CNGrid portal调用GVDS API时，CNGrid用户与GVDS用户一对一映射。

6 虚拟数据空间系统接口实现

系统接口都是基于REST风格的开发API实现的，REST已成为最主要的Web服务设计模型。RESTful Web service是一种轻量级的Web service架构风格，可完全通过HTTP实现，还可以利用缓存提高响应速度，在性能、效率和易用性方面有很好的表现。REST API提供了基于HTTP的国家高性能计算环境访问接口，包括用户管理、作业管理、文件管理等基本功能，也提供了账号管理等高级功能。该接口具有良好的跨语言特性和跨平台特性，使得开发人员可自由选择编程语言。

系统接口在高性能计算环境中为虚拟数据空间系统提供了统一的全局虚拟数据空间的数据访问服务。

- 如图9所示，vlu命令用于获取虚拟数据空间用户的信息。
- 如图10所示，vlg命令用于获取虚拟

命令：*vlu -GVDSAccountName admin -GVDSPassword password*

结果：


```
[
  {
        "GVDSAccountName":"lhj123",
        "GVDSPassword":"lhj123",
        "GVDSAccountID":"0bf36df1-5e3f-4f7e-a251-ec6c8042f4db",
        "AccountEMAIL":"12345678@qq.com",
        "AccountPHONE":"Beijing",
        "AccountAddress":"zky",
        "Department":"false",
        "isroot":"%!s(MISSING)",
        "status":1
  },
  {
        "GVDSAccountName":"lzy",
        "GVDSPassword":"zky",
        "GVDSAccountID":"1",
        "AccountEMAIL":"12345678@qq.com",
        "AccountPHONE":"Beijing",
        "AccountAddress":"zky",
        "Department":"false",
        "isroot":"%!s(MISSING)",
        "status":-402743312
  },
  ......
]
```


图 9　虚拟数据空间用户的信息

命令：*vlg -GVDSAccountName admin -GVDSPassword password*

结果：


```
[
  {
        "UUID":"04541298-7ba5-499f-b3d1-2fb6a554a7bd",
        "name":"mzone1",
        "owner":"mtest",
        "members":["test"],
        "spaces":["f132c571-cd6a-4edf-bb0d-11ae905228dc"],
        "memberauth":{},
        "spaceinfo":[
          {
                "UUID":"f132c571-cd6a-4edf-bb0d-11ae905228dc",
                "name":"mspace",
                "capcity":10,
                "SC_UUID":"3",
                "Storage_UUID":"STOR-3",
                "root_location":"f132c571-cd6a-4edf-bb0d-11ae905228dc",
                "hostCenterName":"Jinan",
                "storageSrcName":"local1",
                "Status":true,
                "replica_spaces":[]
          }
        ]
  }
]
```


图 10　虚拟数据空间用户组的信息

数据空间用户组的信息。

- 如图11所示，vmkuser命令用于创建虚拟数据空间用户。
- 如图12所示，vmoduser命令用于修改虚拟数据空间用户。
- 如图13所示，vscheduling命令可用于获取高性能计算环境信息。
- 如图14所示，vscheduling命令也可用于获取高性能计算环境的队列信息。
- 如图15所示，vscheduling命令还可用于获取高性能环境的应用信息。

环境用户在单个节点就可通过虚拟数据空间系统访问其分布在各个超级计算中心的数据资源（如同访问单一节点的数据一样），并能对这些数据资源进行相应的操作，如编辑、复制、备份等。在任何一个部署了虚拟数据空间系统的节点上，像使用普通Linux文件系统一样，用户能看到其在各个超级计算节点的数据目录和文件，通过虚拟数据空间系统的命令能远程地使用和编辑文件。从用户的角度来看，在虚拟数据空间系统中对文件的操作如同在Linux系统中操作本地文件一样。高性能计算环境通过调用虚拟数据空间系统的相关命令接口实现网格环境中数据文件的操作：如在不同超级计算中心的副本上建立数据文件，在不同超级计算节点上访问/传输文件等。

7 结束语

将虚拟数据空间系统部署到国家高性能计算环境中，可实现虚拟数据空间系统与国家高性能计算环境的深度融合。将虚拟数据空间系统提供的功能补充到国家高性能计算环境中，为国家高性能计算环境提供统一的数据访问视图，方便国家高性能计算环境的用户使用。利用虚拟数

据空间系统提供的数据存储的统一视图，并通过虚拟数据空间系统提供的数据共享等功能实现计算过程中的数据交换方式，可有效地解决高性能计算环境中的计算资源聚合、数据资源分散问题，对国家高性能计算环境的易用性方面起到良好的推动作用。

参考文献:

[1] 吴璨, 王小宁, 肖海力, 等. 基于消息总线的高性能计算环境系统软件优化设计与实现[J]. 高技术通讯, 2020, 30(3): 248–258.
WU C, WANG X N, XIAO H L, et al. Design and implement of middleware for high performance computing environment based on message bus[J]. Chinese High Technology Letters, 2020, 30(3): 248–258.

[2] 王晓东, 赵一宁, 肖海力, 等. 多节点系统异常日志流量模式检测方法[J]. 软件学报, 2020, 31(10): 3295–3308.
WANG X D, ZHAO Y N, XIAO H L, et al. Multi-node system abnormal log flow mode detection method[J]. Journal of Software, 2020, 31(10): 3295–3308.

[3] 和荣, 王小宁, 肖海力. 高性能计算资源聚合服务在中国科技云门户的快速集成研究与实现[J]. 科研信息化技术与应用, 2019, 10(3): 71–78.
HE R, WANG X N, XIAO H L. Research and implementation of high performance computing resource aggregation service in CSTCloud portal[J]. e-Science Technology & Application, 2019, 10(3): 71–78.

[4] 吴燦, 王小宁, 肖海力, 等. 分布式消息系统研究综述[J]. 计算机科学, 2019, 46(z1): 1–5, 34.
WU C, WANG X N, XIAO H L, et al. Survey on distributed message system[J]. Computer Science, 2019, 46(z1): 1–5, 34.

[5] 吴燦, 王小宁, 肖海力, 等. 高性能计算环境中间件的优化设计与实现[J]. 计算机应用研究, 2019, 36(1): 163–167.
WU C, WANG X N, XIAO H L, et al.

命令：

```
vmkuser \
  -GVDSAccountName ljz \
  -GVDSPassword ljz \
  -GVDSAccountID 20201127 \
  -AccountEMAIL 764359986@qq.com \
  -AccountPHONE 19800366030 \
  -AccountAddress Hefei \
  -Department zky \
  -isroot false
```

结果：

```
Dear:ljz, registration success
```

图 11 创建虚拟数据空间用户

命令：

```
vmoduser \
  -GVDSAccountName ljz \
  -GVDSPassword ljz \
  -GVDSAccountID 20201127 \
  -AccountEMAIL 764359986@qq.com \
  -AccountPHONE 13260009966 \
  -AccountAddress Hefei \
  -Department zky \
  -isroot false
```

结果：

```
Modify success
```

图 12 修改虚拟数据空间用户

命令：`vscheduling-query hpc`

结果：

```
Start to query CNGrid...
{
    "type": "hpcRetData",
    "status_code": 0,
    "status_reason": "success",
    "hpcs_list": [
            {
               "city": "BEIJING",
               "hostname": "BeiJing_thu",
               "hpcname": "tsinghua",
               "type": "BRANCH",
               "utime": 1501664783
            },
            {
               "city": "DALIAN",
               "hostname": "DaLian",
               "hpcname": "dicp",
               "type": "BRANCH",
               "utime": 1599358803
            },
            ......
      ]
}
Your query has finished.
```

图 13 获取 HPC 信息

命令：vscheduling -query queue -host era -app vasp

结果：

```
Start to query CNGrid...
{
   "status_code": 0,
   "status_reason": "success",
   "apps_list": [
            {
               "appName": "VASP",
               "appVersion": "",
               "canused": 4,
               "hostname": "HuaiRou",
               "hpcname": "era",
               "maxcpus": 1200,
               "mincpus": 24,
               "njobs": 7228,
               "pendjobs": 160,
               "queuename": "cpuII",
               "runjobs": 7068,
               "walltimelimit": 1450
            },
            {
               "appName": "VASP",
               "appVersion": "",
               "canused": 4,
               "hostname": "HuaiRou",
               "hpcname": "era",
               "maxcpus": 100,
               "mincpus": 8,
               "njobs": 0,
               "pendjobs": 0,
               "queuename": "cpu_dbg",
               "runjobs": 0,
               "walltimelimit": 30
            },
      ......
   ]
}
Your query has finished.
```

图 14 获取 HPC 队列信息

命令：vscheduling -query app -host qsc

结果：

```
Start to query CNGrid...
{
   "status_code": 0,
   "status_reason": "success",
   "apps_list": [
      {
            "appName": "DL_POLY",
            "appVersion": "",
            "canused": 0,
            "hostname": "",
            "hpcname": "",
            "maxcpus": 0,
            "mincpus": 0,
            "njobs": 0,
            "pendjobs": 0,
            "queuename": "",
            "runjobs": 0,
            "walltimelimit": 0
      },
      {
            "appName": "DL_POLY",
            "appVersion": "2.20",
            "canused": 0,
            "hostname": "",
            "hpcname": "",
            "maxcpus": 0,
            "mincpus": 0,
            "njobs": 0,
            "pendjobs": 0,
            "queuename": "",
            "runjobs": 0,
            "walltimelimit": 0
      }
      ......
   ]
}

Your query has finished.
```

图 15 获取 HPC 应用信息

Design and implementation of high performance computing environment middleware[J]. Application Research of Computers, 2019, 36(1): 163-167.

[6] 迟学斌. 高性能计算环境与应用[J]. 国防科技工业, 2018 (5): 21-22.
CHI X B. High performance computing enviromment and applications[J]. Defence Science & Technologg Industry, 2018(5): 21-22.

[7] 迟学斌. 国家高性能计算环境发展报告(2002—2017年)[M]. 北京: 科学出版社, 2018.
CHI X B. Development report on national high performance computing environment(2002—2017)[M]. Beijing: China Science Publishing & Media Ltd., 2018.

[8] 和荣, 王小宁, 卢莎莎, 等. 高性能计算环境通用计算平台[J]. 计算机系统应用, 2019, 28(12): 55-62.
HE R, WANG X N, LU S S, et al. Platform for high performance computing environment[J]. Computer Systems & Applications, 2019, 28(12): 55-62.

[9] 胡正丁, 薛巍. 面向异构众核超级计算机的大规模稀疏计算性能优化研究[J]. 大数据, 2020, 6(4): 40-55.
HU Z D, XUE W. Research on performance optimization for large-scale sparse computation over many-core heterogenous supercomputer[J]. Big Data Research, 2020, 6(4): 40-55.

[10] BIRD I G. LHC computing (WLCG): past, present, and future[J]. Grid and Cloud Computing: Concepts and Practical Applications, 2016, 192: 1.

[11] FOATER I, KESSELMAN C. The grid: blueprint for a new computing infrastructure[M]. San Francisco: Morgan Kaufmann, 2004.

[12] CHOLIA S, SKINNER D, BOVERHOF J. NEWT: a RESTful service for building high performance computing web applications[C]// 2010 Gateway Computing Environments Workshop(GCE). Piscataway: IEEE Press, 2010.

[13] 汤小春, 符莹, 丁朝, 等. 数据流计算环境下

的集群资源管理技术[J]. 大数据, 2020, 6(3): 87–100.
TANG X C, FU Y, DING Z, et al. State-of-art research of cluster resource management in dataflow computing model[J]. Big Data Research, 2020, 6(3): 87–100.

作者简介

何小雨（1990– ），女，中国科学院计算机网络信息中心高性能计算技术与应用发展部、中国科学院大学博士生，主要研究方向为高性能计算、机器学习等。

邓笋根（1975– ），男，中国科学院计算机网络信息中心高性能计算技术与应用发展部高级工程师，主要研究方向为高性能计算软件、算法分析等。

栾海晶（1996– ），女，中国科学院计算机网络信息中心高性能计算技术与应用发展部、中国科学院大学硕士生，主要研究方向为高性能计算和深度学习。

牛北方（1978– ），男，博士，中国科学院计算机网络信息中心研究员，中国科学院大学岗位教授、博士生导师。中国计算机学会高性能计算专业委员会委员。主要研究方向为高性能计算、数据分析算法与软件技术。

收稿日期： 2021-01-21
通信作者： 牛北方，niubf@cnic.cn
基金项目： 国家重点研发计划资助项目（No.2018YFB0203903）
Foundation Item: The National Key Research and Development Program of China (No.2018YFB0203903)