

SKA数据中心云平台 方案的概念性研究

王玲玲, 劳保强, 陆扬, 伍筱聪, 郭绍光
中国科学院上海天文台, 上海 200030

摘要

针对平方公里阵列(SKA)望远镜数据中心海量数据存储和管理的要求,结合当前云计算和大数据技术的发展趋势,对SKA数据中心的云平台方案进行了研究。调研了有关业界的发展状况,对建设数据中心“私有云”进行了前期的一系列尝试,将具体的天文处理软件部署到云平台并进行初步测试。

关键词

平方公里阵列;云计算;数据中心;私有云

中图分类号:TP391

文献标识码:A

doi: 10.11959/j.issn.2096-0271.2016068

Conceptual research of the cloud platform scheme for SKA data centres

WANG Lingling, LAO Baoqiang, LU Yang, WU Xiaocong, GUO Shaoguang
Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China

Abstract

The cloud platform concept of the data center of the square kilometre array (SKA) was investigated, aiming at the big data storage and management requirements of the data centres of SKA. Based on some research of the industry development, the preliminary tests of the construction of data center "private cloud" were presented here, including a specific astronomical processing software deployment and operation in the cloud platform.

Key words

square kilometre array, cloud computing, data center, private cloud

1 引言

平方公里阵列 (square kilometer array, SKA) 望远镜项目是国际天文界自20世纪90年代就开始计划建造的超级望远镜,平方公里是指望远镜阵列的综合接收面积可达一平方公里,整个望远镜最终将由近3 000个抛物面天线和上百万个螺旋阵子天线组成^①。

①

<http://www.skatelescope.org/>

SKA项目涉及除了天文之外的众多领域,集成了各领域的最高科技成果,这需要非常广泛的国际合作才能实现,历经多年的多番论证和谈判,目前确定的主要参加国包括英国、澳大利亚、南非、中国、意大利、荷兰、新西兰、瑞典、加拿大等,还有更多国家和地区在关注或洽谈中。项目计划从2018年开始建设,到2030年结束,共分两个阶段进行,即所谓的SKA1(2018—2023年)和SKA2(2024—2030年)^②,届时将分别在南非和澳大利亚两个站址国建设数据处理中心。

②

<http://china.skatelescope.org/>

③

<http://www.nscg-gz.cn>

2 面临的挑战

2.1 大数据量

SKA的数据量可以说是惊人的,仅在SKA1阶段,需要长期存档的科学数据就将每年增加50~300 PB,到了SKA2阶段,这个数字将提高100倍,庞大的数据量和预算的限制决定了所有数据都由站址国数据处理中心分析处理是不可能的,还需要在其他国家或地区建设区域数据中心来进行相应的数据处理,并更好地为区域内科学家开展科学研究提供必要的计算和存储资源,为此,SKA总部已明确提出了建设

区域中心的要求,并组建了协调小组负责区域中心的顶层设计。区域数据中心的数据处理量或许会少于站址国数据处理中心,但在量级上是相当的,具体的规模和设计取决于各承建国的预算、需求和能力。

SKA数据中心是一个共享的平台,需要承担科学分析、数据存储和备份的任务,需要为科学家提供便捷、高效的数据处理和分析的服务,按照SKA的数据量和实时处理要求,这势必要依赖超级计算机的支持,但若是按照现有架构设计,不仅投资巨大,光是设备运转的电费成本都将是一个天文数字。因此,设计低能耗、低成本、更先进的计算架构是SKA数据中心建设迫切需要解决的一大难题。

2.2 高性能计算和云计算

高性能计算向来是密集型科学计算的首选,多采用昂贵的服务器和顶尖的硬件配置来实现,这其中最著名的当属连续6年位居世界第一的“天河二号”^③。然而,如前所述,这些超级计算机的建设和运营成本太高,若要达到SKA的数据存储和实时处理要求,需要的造价要远远超出预算。

近几年来,云计算已经越来越得到业界青睐,以“亚马逊云”为代表的一大批云服务已向公众开放,企业用户纷纷向各种“云”靠拢,各种云计算技术层出不穷,可以说云计算正以一种“不可替代”“不可阻挡”的趋势发展。云计算的本质就是将众多普通计算机通过网络连接,共同解决大数据量的计算问题,是一种较低成本的、可扩展的计算架构。

事实上,国际同行已经分别对云、超算和集群3种平台下的软件运行进行了性能测试,结果表明云计算的表现还是

很乐观的,尤其在针对不同的计算实现灵活调配硬件资源方面具有明显的领先优势^[3]。

针对SKA的实际情况和高性能计算、云计算的各自优势,笔者认为高性能计算和云计算两者的融合将很可能是一种值得探讨的方案。在本文中,笔者将对若干云平台进行一些尝试,包括将天文数据处理软件部署在云平台上运行。

2.3 天文数据处理软件的瓶颈

几十年的天文科学发展,产生了众多的天文数据处理软件,甚至针对某些同一类的科学观测内容,也会有多种软件可供科学家选择,鉴于SKA项目的复杂性和长期性,将考虑采用现有的主流射电天文数据处理软件,继承其主要架构和库并针对SKA的数据特点进行修改更新。其中,大型天文数据处理软件普通天文软件应用(common astronomy software applications, CASA)是目前SKA采用的数据分析软件,它是以美国国家射电天文台(national radio astronomical observatory, NRAO)为首的众多国际天文界的科学家共同开发完成,CASA不但可以处理干涉阵的数据,也可以进行单天线数据的处理,目前已广泛应用于射电天文界^④。然而,CASA的原初设计并没有考虑到PB量级的大数据,因此它是基于单点串行的,并不能很好地支持并行,这一点局限了CASA在未来大规模天文数据处理中的使用。为了解决这一难题,国际同行已在其数据I/O接口的改造方面做了很重要的工作,在一定程度上提升了数据读写速度^[4]。

笔者以CASA为例,将其部署到云平台上运行,研究在云平台上提高CASA性能的方法。

3 云平台的选择

云计算的种类很多,根据数据公开的程度可以分为公有云、私有云、混合云;而从技术角度或者服务层面又可以分为基础设施即服务(infrastructure as a service, IaaS)、平台即服务(platform as a service, PaaS)和软件即服务(software as a service, SaaS)。用户有两种选择:使用成熟的公有云服务或是自己搭建私有云。笔者选取了其中有代表性的产品逐一进行了尝试。

3.1 公有云

在公有云领域,亚马逊云、Microsoft Azure、阿里云等都是业内著名产品,厂商提供服务,用户租用所需资源或部署应用、开发产品,用户不需关心底层运维的一切事务,可大大节省精力去关注产品或应用本身。这种云都具有复杂的计费系统,根据用户使用资源的大小、数量、时长等进行收费,或者有若干免费项目可供用户试用。笔者在亚马逊云上进行了尝试,注册后登录,选择区域(需要注意的是每个区域能够提供的服务会有些许差别),在控制页面可以直观地看到众多服务,可以看出,亚马逊云虽然是国际上IaaS市场占有率最大的厂商,但是它的服务范围并不限于IaaS的资源,还提供了很多PaaS及SaaS的服务和工具可供选择,这些大大提升了用户体验^⑤。

这里主要关注其中的亚马逊弹性计算云(Amazon elastic compute cloud, EC2)服务,该服务提供在云中可调整大小的计算容量,即可以提供任意大小的虚拟服务器。使用时,首先,启动一个实例,

④ <http://casa.nrao.edu/>

⑤ <http://aws.amazon.com/>

也就是启动一个虚拟服务器,选择一个镜像,按照提示一步步配置,创建密钥对并下载,完成。点击查看实例可以看到新启动的实例的详细信息,使用安全外壳协议(security shell protocol, SSH)连接实例(连接需要验证之前创建的密钥对)进行操作。

由于亚马逊云并没有在中国建立数据中心,笔者测试时使用的区域是美国东部(美国弗吉尼亚州北部),所以连接实例时有时会产生命令延迟的感觉,即使尝试亚太区域(日本东京)也没有明显提高。

3.2 私有云

“私有云”顾名思义就是可供一定范围内的用户使用的专有云,用户自己拥有基础设施,自行安装专用的软件,同时也需自行控制网络和数据的安全性。基础设施架构的选择有VMware和OpenStack等,其中后者是开源的,鉴于研究经费限制,这里重点关注后者。

OpenStack由美国国家航空航天局(National Aeronautics and Space Administration, NASA)和Rackspace(全球三大云计算中心之一)共同发起,是如今最流行的大型开源项目,是全球开发人员的合作成果,旨在提供标准的公有云和私有云的云计算平台,背后有着活跃的开发社区和一些业界最大企业的支持,其社区是全世界增长最快的开源社区之一,目前已有59 583人,涉及612个企业、184个国家和地区,共产生了超过2 000万行代码^⑥。

OpenStack本质是一个云操作系统,控制整个数据中心的大量计算、存储和网络资源,所有这些都通过仪表盘管理,管理员授权用户通过Web界面或者应用程序编程接口(application programming

interface, API)访问资源,在这一点上,所有的基础设施架构都差不多。

再进一步细分,OpenStack由多个针对不同用途的项目组成,比如针对计算、块存储、对象存储、镜像、网络、认证等服务,分别称为NOVA、CINDER、SWIFT、GLANCE、NEUTRON、KEYSTONE等,这些称之为核心服务,用户可以根据需求选择安装与否。

4 测试

本节重点介绍OpenStack私有云的搭建和CASA软件在其中运行的体验过程。

4.1 安装OpenStack

OpenStack的安装方法有多种,第一种是直接下载源代码安装。按照文档介绍一步步配置网络,安装各组件,这是最直接的一种安装方式,但是这对于初学者来说太过复杂,门槛较高。尤其是OpenStack的网络很复杂,标准的配置有4个不同的网络层:管理网络、数据网络、外部公共网络和API网络,安装人员必须对网络尤其是虚拟网络技术有很深刻的理解,并且需在进行一定的规划和设计之后才能进行,否则就需要全部重新开始。

第二种方法是官网推荐的使用DevStack快速安装。DevStack是一系列脚本,可以帮助用户在单机上很快速地安装完整的OpenStack实验环境,整个安装过程非常简单,笔者测试用的计算机的操作系统是Ubuntu14.04,按照说明逐步进行操作,具体如图1所示。

第三种安装方法是利用业内厂商集成的软件来安装。众多专注于OpenStack的企业提供了各具特色的集成服务,比如

^⑥
<http://www.openstack.org/>

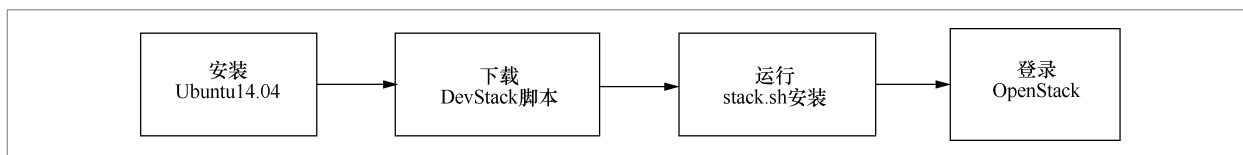


图1 DevStack 安装示意

IBM、CISCO、RedHat、Mirantis、九州云、海云捷迅、EasyStack、UnitedStack等。这几家厂商的产品各有侧重，最终，笔者选用了业内普遍好评的Mirantis公司的Fuel工具来试用。

Fuel是专门用于部署和管理OpenStack的实用工具，部署过程中可以任意选择OpenStack版本或是其他第三方插件，并实时查看部署状态，部署完成后可以随时进行维护或扩展自己的云。另外，Fuel还可以轻松实现部署并管理多个云，所有这些操作均通过Web界面进行，用户体验好^⑦。

笔者的测试在一台运行Windows7系统的笔记本上进行，使用VirtualBox工具生成多个虚拟机进行安装。详细过程如图2所示，过程中全部选用默认设置。

4.2 安装CASA软件并测试运行

目前CASA的最新版本是v4.7.0，笔者测试时用的版本是v4.5.3，因为创建的实例

是Ubuntu系统，所以按照推荐，选择了其中的Red Hat 6版本，安装过程跟本地安装完全相同。

CASA软件的数据处理过程由很多任务和工具组成，用户可根据数据情况选择使用，主要的任务包括Manipulation、Calibration及Imaging等^⑧。

实际测试过程选用了一段大小为596 MB的数据，数据来自坐落于智利的阿塔卡玛毫米/亚毫米波阵列(atacama large millimeter/submillimeter array, ALMA)望远镜，观测时间近2 h，已经过校准。创建存储卷并挂载到该实例，运行CASA执行clean(清除)任务，运行过程跟在本地执行任务相同。

然后，将此实例制作映像，以此为准再次创建并启动5个实例，分别运行split(拆分)及clean任务，其中split任务以数据的每个field(区域)为范围进行，中间数据及结果数据复制到存储卷备用。具体测试流程如图3所示。

测试过程中各实例各任务执行所花费

⑦

<https://www.mirantis.com/>

⑧

<https://casa.nrao.edu/docs/UserMan/index.html>

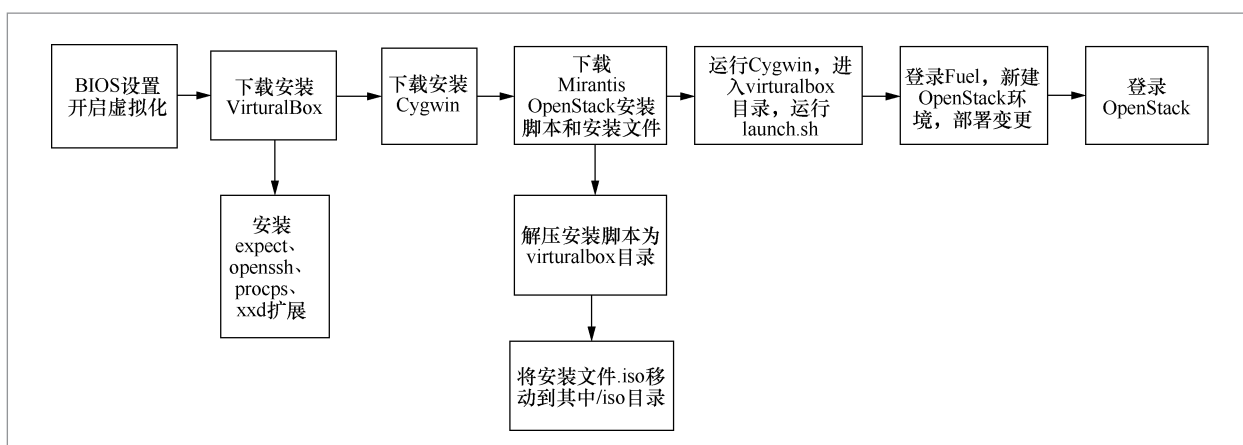


图2 使用 Fuel 安装 OpenStack 示意

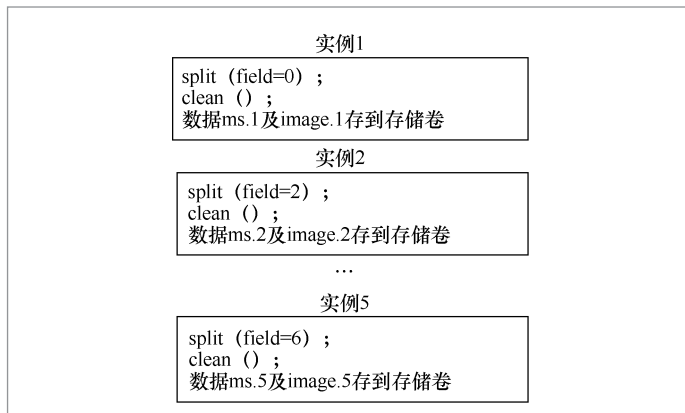


图3 启动5个实例同时开始处理数据

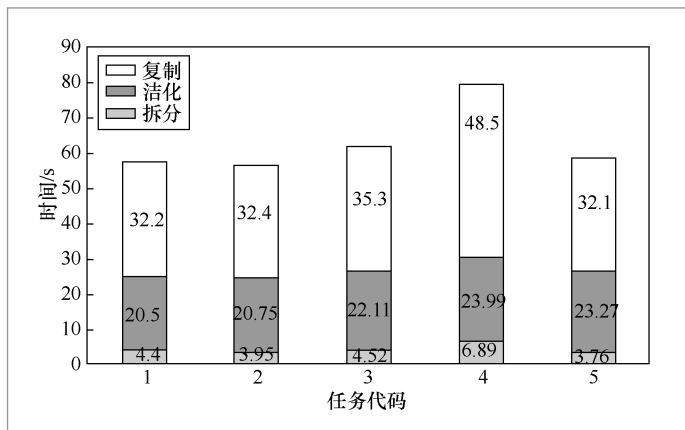


图4 执行各个任务所花费的时间

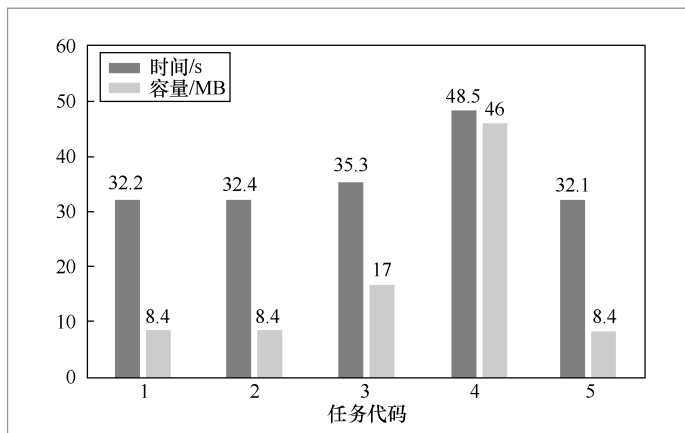


图5 数据容量与复制时间示意

⑨
<http://docs.openstack.org/developer/language-bindings.html>

的时间统计如图4所示,因为5个实例同时运行,所以其中耗时最多的实例所用时间可视为整个任务的执行时间,显而易见,其中实例4总用时最多,达79.38 s,而同样

的数据若由一台主机来处理,各任务加起来则需用时180.5 s,所以,云平台的处理速度优势很明显。

仔细分析可发现,在各个实例的任务执行过程中,耗时最多的任务是数据复制,即数据从实例复制到公用存储卷所花费的时间最多。不过,随着数据的增长,数据I/O口对整个数据处理时间的影响会大大削弱(如图5、图6所示)。

4.3 讨论

上述测试中,创建和运行的实例是通过仪表盘Web应用程序来手动执行的,如果希望实现自动化运行应用,则需要编写自动化脚本,而由于所有OpenStack服务都公开了原生的Python API绑定,所以用户只要安装OpenStack客户端就可以进行自动化配置及操作^⑨。具体使用方法是先安装相应服务的工具,比如若要使用计算服务,则先安装novaclient,若要启动一个实例,则只需一条指令。

上述只是一个演示性测试,若真正用于天文数据处理,还有很多细节问题需要考虑,以目前的结果来看,效果已经很显著,可以说,云平台方案对于天文数据中心的的应用前景是乐观的。同时,云平台的建设并不需要特别昂贵的基础设施成本投入,这对于一个数据中心的建设来说无疑是个很大的优势。

另外,值得注意的是,由于将来要处理的数据是海量的,所以在实际运行中需将数据尽量小地分割,尽量降低单个实例上运行的数据量,最后再将数据整合。根据天文数据的特点,可以按照频率、通道或时间来切分。具体实施需要继续测试,如图7所示。

分割的过程中,需注意数据分割的粒度问题,即仔细分析每种实例一次处理多

大的数据是最优化的, 依此来确定在开始时需将数据分割成多大。另一个很重要的问题是资源调度, 实际运行中每个实例的处理速度不完全相同, 如何给实例分配资源、创建启动的顺序从而实现协同处理, 也是需要仔细研究的课题。

5 结束语

由于资源的限制, 本次原理性验证实验只是构建了一个最基本配置的测试平台, 节点少, 配置低, 使用的测试数据量也不太大, 目前的工作只是万里长征的第一步, 进一步的大规模性能测试需要扩展云平台的资源后进行。在后续研究中, 将综合比较云平台与超算平台下软件的执行效率等性能, 详细研究SKA数据中心的建设方案。

参考文献:

[1] DEWDNEY P E, HALL P J, SCHILLIZZI R T, et al. The square kilometre array[J]. Proceedings of the IEEE, 2009(97): 1482-1496.
 [2] 彭勃, 金乘进, 杜彪, 等. 持续参与世界最大综合孔径望远镜SKA国际合作[J]. 中国科学: 物理学力学天文学, 2012, 42(12): 1292-1307.

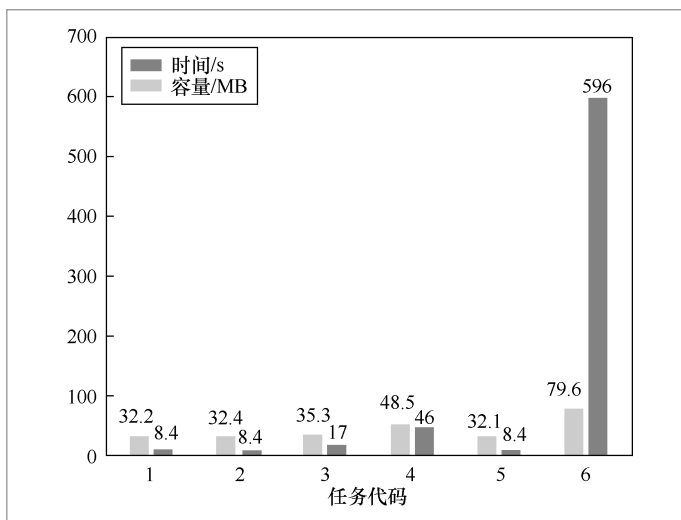


图6 数据容量与复制时间示意(与原始数据比较)

PENG B, JIN C J, DU B, et al. China's participation in the SKA-the world's largest synthesis radio telescope[J]. Scientia Sinica: Physica, Mechanica & Astronmica, 2012, 42(12): 1292-1307.

[3] DODSON R, VINSEN K, WU C, et al. Imaging SKA-scale data in three different computing environments[J]. Physics, 2015(14): 8-22.
 [4] WANG R, HARRIS C, WICENEC A. AdiosStMan: parallelizing casacore table data system using adaptive IO system[J]. Astronomy and Computing, 2016(16): 146-154.

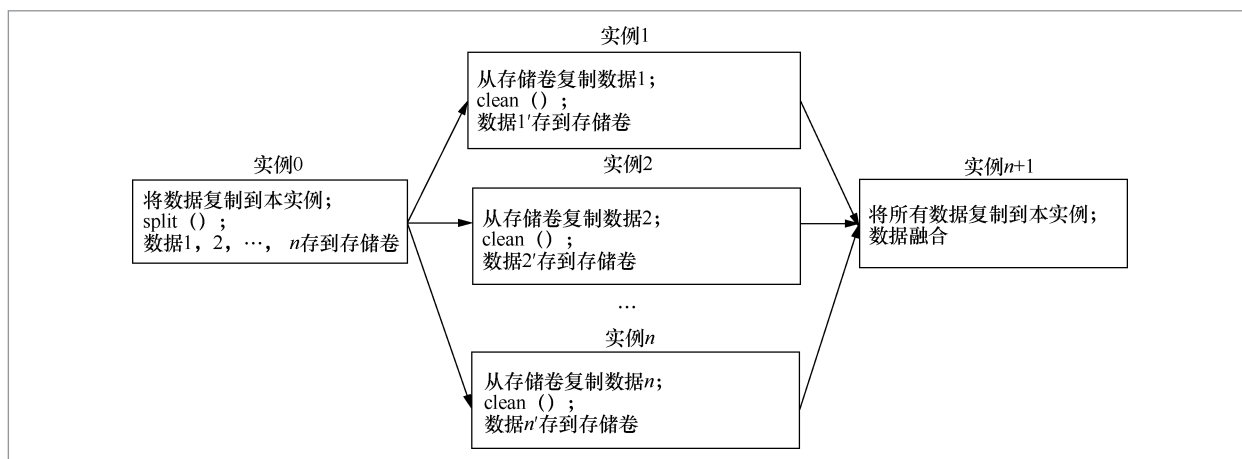
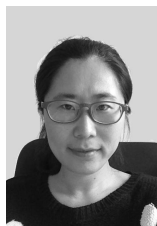


图7 启动批量实例协同处理数据示意

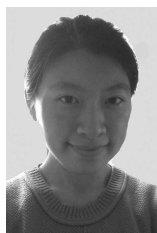
作者简介



王玲玲 (1978-), 女, 中国科学院上海天文台高级工程师, 主要研究方向为射电望远镜台站监控、台站时频系统、天文大数据存储和云计算等。



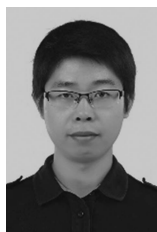
劳保强 (1989-), 男, 中国科学院上海天文台助理工程师, 主要研究方向为存储底层I/O并行技术和数字波束形成技术。



陆扬 (1985-), 女, 中国科学院上海天文台助理工程师, 主要研究方向为天文与统计。



伍筱聪 (1989-), 女, 中国科学院上海天文台助理工程师, 主要研究方向为SKA数据处理。



郭绍光 (1985-), 男, 中国科学院上海天文台工程师, 主要研究方向为VLBI技术、相关处理及终端存储。

收稿日期: 2016-10-08

基金项目: 中国科学院天文财政专项“SKA1科学与数据中心概念性研究”基金资助项目

Foundation Item: Conceptual Studies of the SKA-1 Science and Data Centre, Special Funding for Astronomy of the Chinese Academy of Sciences