

· 应用 / APPLICATION ·

# 一种基于 Linux 容器技术的大规模遥感数据云服务平台

张品<sup>1,2</sup>, 张海明<sup>1</sup>, 黎建辉<sup>1</sup>

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

**摘要:** 遥感数据云服务平台是基于云计算技术, 整合大规模遥感数据的存储资源和计算资源, 实现资源共享和按需使用的服务模式的地理信息处理服务平台。我们基于分布式存储技术实现遥感数据的高效存储、Linux 容器技术实现快速部署和资源隔离、ownCloud 私有云技术实现高效共享和 IPython notebook 交互式技术实现方便易用交互, 设计了一种稳定、高效的地理信息云服务平台。用户可通过 Web 的方式方便的访问大规模遥感数据, 并利用云主机的计算、存储资源对所需的遥感数据进行分析 and 处理。

**关键词:** 遥感图像; 地理信息; 云服务; 云计算; 容器技术

doi:10.11871/j.issn.1674-9480.2015.02.006

## A Large-Scale Remote Sensing Data Cloud Service Platform Based on Linux Container

Zhang Pin<sup>1,2</sup>, Zhang Haiming<sup>1</sup>, Li Jianhui<sup>1</sup>

1. Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China

2. University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract:** Remote sensing data cloud service platform is the geographic information processing service platform based on cloud computing technology, integrating computing resources and storage resources of large-

---

基金项目: “十二五”国家科技支撑计划项目(2013BAD15B02); 国家自然科学基金(91224006); 中国科学院战略性先导科技专项项目(XDA06010307, XDA05050601, XDA05150401)

scale remote sensing data, to achieve the patterns of resource sharing and on-demand services. This paper present a platform which is implemented based on distributed storage technology, Linux container technology, ownCloud private cloud technology and IPython notebook interactive technology. It is easy to access the large scale remote sensing data through web as well as to analyze and process the data using the computation and storage resources on the cloud host.

**Keywords:** remote sensing image; geographic information; cloud service; cloud computing; container technology

## 0 引言

遥感一词来源于英语“Remote Sensing”，直译为“遥远的感知”，后来人们将它简译为遥感。遥感广义上是指一切无接触的远距离探测，包括对电磁场、力场、机械波（声波、地震波）等的探测，狭义上是指应用探测器，不与探测目标相接触，从远处把目标的电磁波特性记录下来，通过分析，揭示出物体的特征性质及其变化的综合性探测技术。遥感技术在二十世纪六十年代初发展起来，开始时为航空遥感。

1972年美国发射了第一颗陆地卫星后，航天遥感迅速发展。遥感已经成为一门实用的，先进的空间探测技术。遥感探测能在较短的时间内，从空中乃至宇宙空间对大范围地区进行对地观测，对地球资源和环境分析极为重要。

遥感具有采集更新快、多分辨率、信息丰富、数据类型多等特点，已经成为地理信息科学的主要信息源。目前，遥感数据最常用的处理方法为遥感数据应用人员使用个人计算机或工作站基于商业或者开源地理信息软件对所需数据进行分析处理。其基本流程为：(1)从互联网下载或其他存储介质中复制遥感数据到用户个人计算机；(2)使用遥感分析软件中的相关分析工具或调用遥感分析软件的编程接口编写并运行脚本处理数据；(3)存储结果。随着各类遥感应用发展，所需处理遥感数据的空间覆盖与时间跨度要求越来越大。受制于计算机内存和存储空间的限制，这种模式已经无法满足如此大规模数据的存储和处理需求。面对海量遥感数据，如何对其高效检索、获取、存储和分析是相关技术人员面临的一项极大挑战。

为了解决科研活动中上述问题，本文利用分

布式存储技术和 ownCloud 私有云技术提供存储，Linux 容器技术进行资源隔离和计算处理，并将存储和计算有机结合，设计实现了一种方便、易用、稳定、高效的遥感数据云服务平台。基于 Web 采用 IPython notebook 交互式技术提供给遥感应用人员。该平台不仅可扩展性好、可操作性强，还可以进行远程管理和控制，使遥感应用人员可以高效地检索、获取、存储和分析遥感数据，降低了遥感数据的应用门槛和提高了应用效率。

### 0.1 遥感数据云服务平台研究

随着遥感数据的大量产生和遥感产业的发展，如何处理遥感大数据已经成为不可规避的技术难题。现有的将数据存储到本地，再将数据逐个传递的旧模式明显不适合对大数据的处理。同时，还会增加网络负载，存储设备、数据中心机房、电力供应等都会产生大量的昂贵费用。因此，遥感产业与云计算的结合变得迫切而必要。随着云计算在遥感领域的成熟，将会极大提高数据处理能力，丰富产品种类，降低成本，促进产业内大企业的发展，以及云服务带来的数据共享能力和商业模式服务无疑将促进遥感的产业化发展<sup>[1]</sup>。

遥感数据云服务是基于云计算技术，整合大规模遥感数据资源和技术资源，基于互联网环境以按需共享的方式提供在线遥感应用服务<sup>[2]</sup>。遥感数据云服务具备云计算两方面的优势，一是通过云计算技术提高遥感数据储存和处理的效率，二是实现资源整合、按需共享的服务模式。通过云计算的模式，用户可以利用服务器端(即存储数据的云端)来完成海量数据的访问处理及复杂算法的运算等“重量级”工作，避免终端客户被迫传输或者管理原始数据，

而在移动设备上, 只需基于互联网进行访问控制和算法定义等“轻量级”应用, 不需要耗费太多本地的硬件资源和网络资源。

随着云计算技术逐步应用和推广, 基于互联网环境下的地理空数据网络处理服务逐步受到关注。开放地理信息系统协会 (Open Geospatial Consortium, OGC) 制定了针对地理空间数据的网络处理服务标准 (Web Processing Service, WPS), 规范了网络处理服务的通信协议 (<http://www.opengeospatial.org/standards/wps>)。利用现有采用该协议实现的系统, 由于并没有给用户提供了模型和算法插件接口, 遥感应用人员只能基于系统内置的分析模型和工具, 提交计算任务, 等待系统返回处理结果, 却无法在平台中运行自己设计的分析模型和工具, 不利于用户的个性化需求, 也阻碍了模型和算法的共享。

在遥感产业发展过程中, 遥感数据处理软件 ArcGIS、Grass GIS 等得到广泛适用。在面对海量遥感影像数据的产生和遥感应用人员对遥感数据处理服务需求的日益增长, 正是由于用户对桌面软件的依赖, 现有遥感云服务平台或关键技术研究均为基于管理程序的虚拟化技术。随着用户基数的增长, 这种架构设计很难满足大量用户的访问, 相反硬件资源的需求也会急剧上升。

## 0.2 虚拟化技术研究

虚拟化 (Virtualization) 是指单一物理资源的多个逻辑表示或者多个物理资源的单一逻辑表示, 即资源的抽象化。简单的来说, 是指计算机元件在虚拟的基础上而不是真实的基础上运行。虚拟化技术能够将有限的固定资源根据不同的需求进行重新规划达到最大利用率, 是实现优化资源, 简化管理等目的的解决方案。和传统的 IT 资源分配应用方式相比, 虚拟化技术有以下优势: 虚拟化技术可以大大的提高资源的利用率, 提供相互隔离、安全、高效的应用执行环境, 采用虚拟化技术后, 虚拟化系统能够方便的管理和升级资源<sup>[3]</sup>。

随着虚拟化技术的发展, 市面上已经出现

了很多成熟的虚拟化解决方案和产品, 包括传统的通过使用控制程序 (Control Program, 也称 Virtual Machine Monitor 或 Hypervisor), 隐藏特定计算平台的实际物理特性, 为用户提供抽象的、统一的、模拟的计算环境 (称为虚拟机) 的技术。例如 VMware vSphere、微软 Hyper-V, 开源的 Xen、KVM 等。此外, 基于容器的虚拟化技术 LXC (Linux Containers) 也开始显示出其优势, 慢慢成为企业和用户的选择之一, 例如 Docker<sup>[4-5]</sup>。

基于容器的虚拟化和传统的 VMM 模式的虚拟化相比, 最大的不同在于容器内每个虚拟机安装的不是完整的操作系统实例, 这种技术仅安装每个操作系统实例的一部分, 每个实例以智能的方式与安装在主操作系统内核中的虚拟化层协同工作。LXC 是一种内核虚拟化技术, 可以提供轻量级的虚拟化, 以便隔离进程和资源, 而且不需要提供指令解释机制以及全虚拟化的其他复杂性。所谓轻量级虚拟化, 也指代操作系统级别的虚拟化, 通过内核和用户态进程组的支持, 实现的独立网络 IP、进程树等类似虚拟机的隔离运行环境, 但是和宿主机运行同样的内核。在资源管理方面, 容器依赖 Linux 内核的 cgroups 子系统, cgroups 子系统是 Linux 内核提供的基于进程组的资源管理框架, 可以为特定的进程组限定可以使用的资源; 在资源隔离方面, 容器引入了 Linux 内核的 namespace 特性, 即在资源复制时加入了相应的标记。所以, 容器能够有效地将由单个操作系统管理的资源划分到孤立的组中, 以更好地在孤立的组之间平衡有冲突的资源使用需求。

Docker 是一个构建在 LXC 之上的, 基于进程容器 (Process container) 的轻量级 VM 解决方案。与传统 VM 解决方案相比, Docker 最大的区别是不包含操作系统内核。由于 Docker 轻量级虚拟化的特点, Docker 具有以下典型的应用场景: 构建隔离的标准化运行环境, 轻量级的 PaaS, 构建自动化测试和持续集成环境, 以及一切可以横向扩展的应用。大规模遥感数据云服务平台给用户提供了标准化的遥感数据处理平台, 遥感应用人员只需要设置自定义参数和提交个性化代码即可满足遥感数据的处理需

求。在面对大规模用户访问时，Docker 可以迅速启动并支持硬件资源和软件资源的动态扩展。由此可见，Docker 技术很适合大规模遥感数据云服务平台的设计需求。

## 1 系统总体设计架构与流程

遥感数据云服务平台设计主要分为三个主要部分(图1)：

### (1) 基础数据存储

基于开源分布式文件系统 MFS 和 Linux 集群构建存储环境，对大规模遥感图像进行高性能的分布式存储，基于 Posix 标准文件系统接口的对其进行统一集中的存储与管理。在分布式文件系统中整理存储 LANDSAT、MODIS、NOAA AVHRR 等遥感图像数据文件，并通过云主机作为分布式文件系统客户端挂载数据目录，以 NFS 服务方式导出数据目录进行共享。

## (2) 平台云计算资源管理与分配

基于 Linux 容器虚拟化技术 Docker，提供轻量级的虚拟化方案。Docker 启动快占用资源少，能有效地对云主机计算资源进行隔离，使得云主机的计算资源利用率达到最大化。通过 Docker 主机服务系统，可快速构建 Docker 服务器模板，模板包括用户名密码信息，必要的运行库，系统使用说明，系统文件说明，demo 程序等。

### (3) 用户 Web 访问接口

基于 Nginx 的 IP 和名称的虚拟主机服务, 映射不同的 Docker 地址到 Web 访问入口, 用户通过用户名和密码即可访问不同的 Docker 服务器。在用户访问入口, 通过 IPython notebook 技术, 用户可编写、维护、运行自定义 python 代码访问遥感数据并进行计算, 获取自定义的遥感信息。用户的运行结果可保存到用户共享目录中, 通过 ownCloud 私有云技术对结果文件进行下载, 也可自己上传自定义文件用于计算<sup>[6]</sup>。

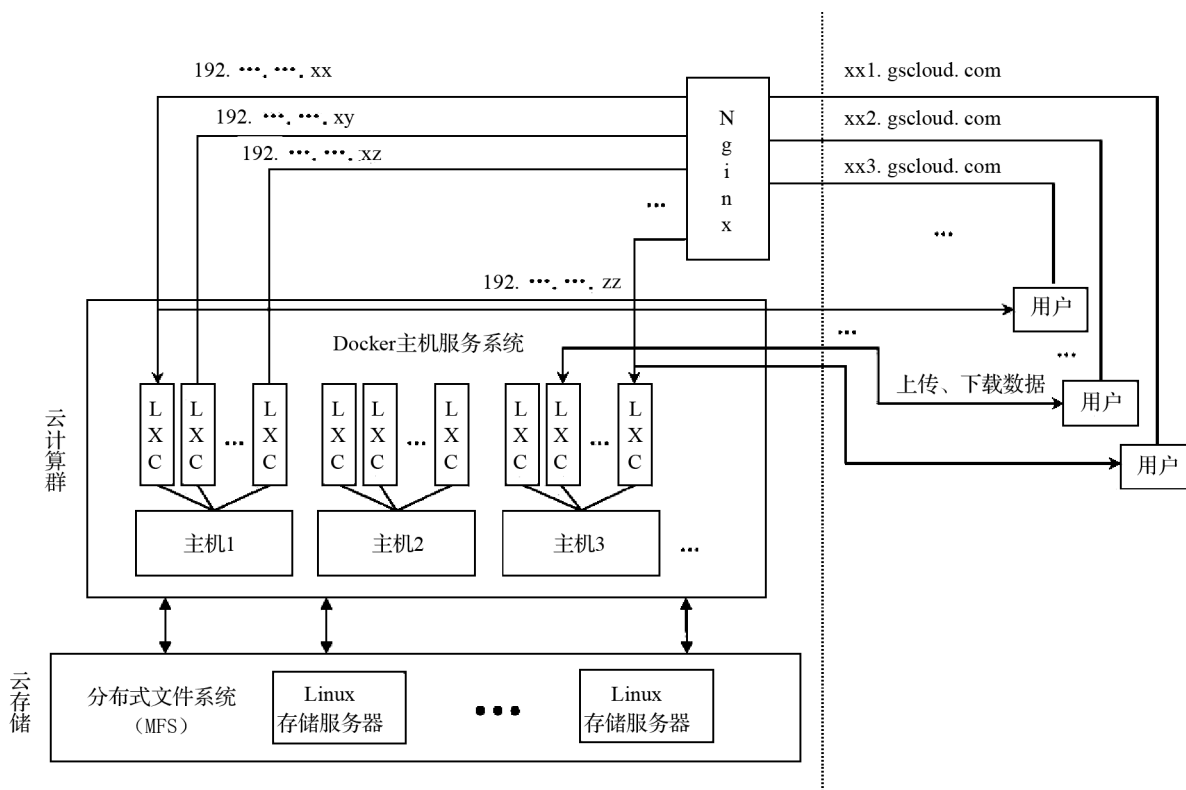


图1 遥感数据云服务平台整体架构

**Fig. 1 The framework of remote sensing data cloud service platform**



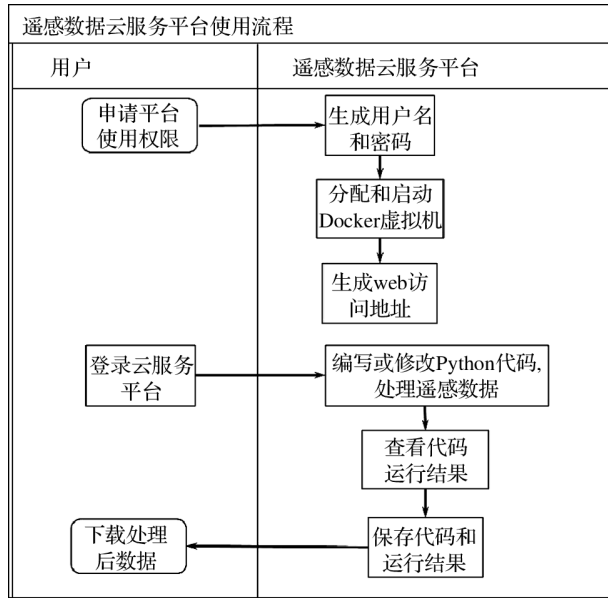


图2 遥感数据云平台使用流程图

Fig. 2 The flow chart of remote sensing data cloud service platform

基于平台设计方案, 整个遥感数据云服务平台使用流程可归纳为三个部分: 资源申请和分配, 资源访问和使用, 数据上传和下载。图2 是实现的一个简化应用流程。

## 2 原型系统实现和关键技术

### 2.1 遥感数据云服务平台原型实现

中国科学院计算机网络信息中心已经具备成规模的存储与计算资源环境, 科学数据中心更是积累了大量的各类遥感图像及相关产品(数百 TB)。基于这些基础资源开发了一个大规模遥感数据云服务平台, 初步实现了遥感数据的存储、计算和大规模访问。图3 是遥感数据在 MFS 文件系统中的存储结构, 目前已收录 300 多 TB 全球 10 年内的遥感图像数据(包括部分成型的遥感数据产品), 图4 是遥感图像示例文件。

整个遥感数据云服务平台的使用流程, 主要包括以下 3 个使用步骤:

(1) 用户通过 Web 界面申请遥感数据云服务平台使用权限, 填写相关信息提交管理员审核。审核通过后, 管理员给用户生成用户名密码, 通

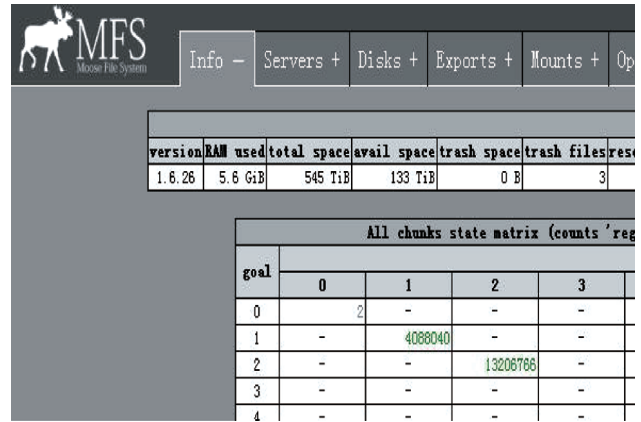


图3 遥感数据存储结构

Fig. 3 Remote sensing data storage structure

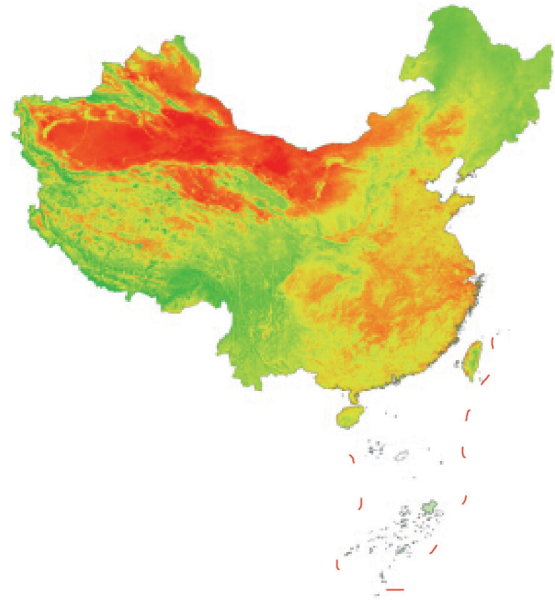


图4 全国范围遥感图像实例

Fig. 4 Nationwide remote sensing image sample

过 Docker 主机服务系统生成 Docker 模板, 并将生成的访问地址, 用户名密码通过邮件形式发送给用户, 完成整个申请流程(图5);

(2) 用户通过 Web 界面使用申请的用户名和密码登录遥感数据云服务平台, 编写或修改 Python 代码读取和计算遥感图像文件(图6 为编写代码读取 2008 年每日遥感图像文件, 采集北京区域每个栅格的昼夜温度数据, 去除空值后取平均得到的全年温度值)。点击运行可查看运行结果, 可保存代码和运

行结果至用户私人目录(图7);

(3) 用户自定义代码运行结束后, 通过 Web 登录遥感数据云服务平台私人目录, 上传和下载文件或运行后处理的数据。

在实际的运营当中, 用户的访问申请和信息审核均通过系统管理员来人工审核。整个遥感数据云服务平台的使用手册, 遥感图像文件说明(存储结构和文件存储信息), demo 运行程序和系统接口功能和参数说明均可以通过 Web 访问, 大大降低了遥感数据的应用门槛和提高了应用效率。用户不需要关注

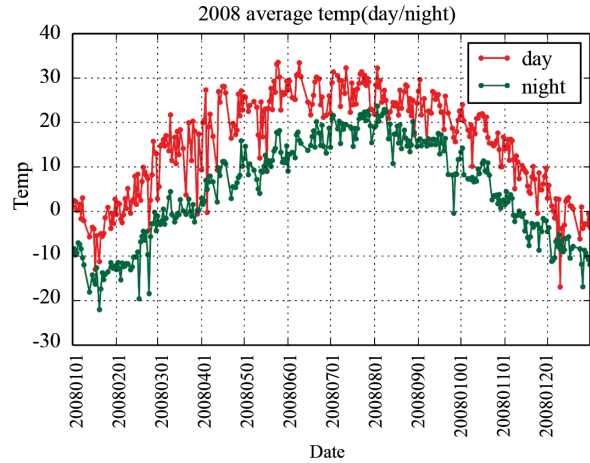


图7 2008年北京日夜间平均温度

Fig. 7 The average temperature at day and night in Beijing, 2008

基于 Linux 环境面向遥感数据的编程环境构建, 在了解 Linux 系统的背景下, 基于系统定义的遥感数据读写接口、通用算法接口以及并行调度管理接口即可实现遥感数据的处理和应用。一方面用户基于遥感数据读写接口和通用算法接口可以运用 GDAL 库对遥感图像进行处理, 另一方面平台还为用户提供了 MPI 并行接口、并行调度管理接口以及计算资源, 使得用户在进行大量遥感图像处理时可以并行计算。相比于串行处理的处理方式, 以计算北京市日夜平均温度为例, 串行处理用时 3 分 31 秒, 16 核并行化后缩减为 24.33 秒, 加速比为 8.67, 大大提高了遥感应用的效率。正是由于基于 Linux 容器虚拟化技术, 整个遥感数据云服务平台可以满足在保证用户计算效率的情况下, 支持大量用户同时访问遥感数据。保证了整个系统的性能和可用性。

由于遥感数据处理既要满足先提交先服务调度 (First-Come-First-Served, FCFS) 的服务机制, 又要满足应急条件下一定的快速响应能力, 在并行调度管理中, 本平台采用了基于优先级和时间戳的调度策略。在用户提交计算任务时, 任务被赋予两个属性: 优先级和时间戳。根据优先级和时间戳对任务进行排序, 优先级较高的任务先调度, 当优先级较低的任务等待时间超过系统设定的阈值时, 则提高一级任务的优先级并重置该任务的时间戳为当前时间。

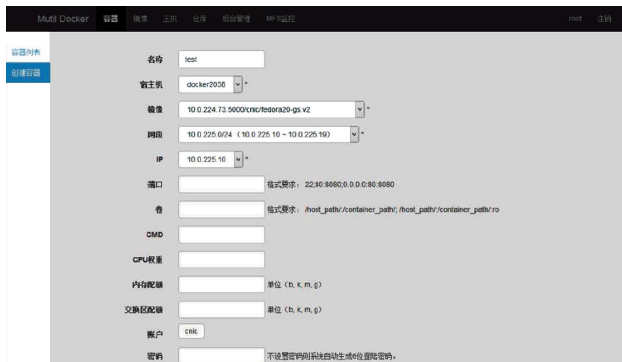


图5 遥感数据云服务平台资源管理与分配

Fig. 5 Resource management and distribution of remote sensing data cloud services platform

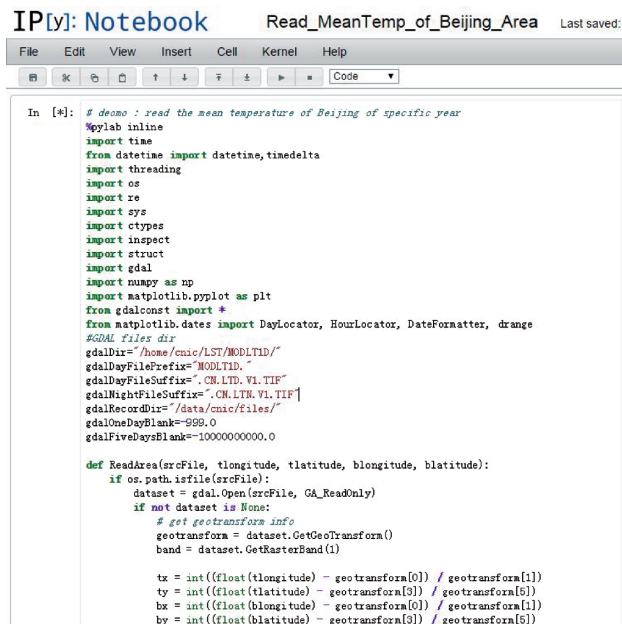


图6 自定义代码读取北京地区每年温度数据

Fig. 6 Custom code to read temperature data every year in Beijing

遥感数据云服务平台实现数据共享的一个关键问题是如何既允许用户自由的使用共享数据，又能防止用户拥有和转移数据。提供遥感数据的下载面临数据安全问題，仅提供数据访问接口往往又难以满足应用需求。这一问题在很长一段时间内严重制约着遥感数据共享和实际的应用推广<sup>[7-8]</sup>。本遥感数据云服务平台为解决这一问题提供了基于用户身份认证机制、用户属性和遥感数据属性的多维细粒度访问控制策略的新的技术方案，完善平台认证机制和增强访问控制的目的。基于云服务平台，用户可以在保证计算效率的情况下自由的使用和处理拥有访问权限的遥感数据图像文件，既限制了遥感数据的流失又保证了用户的个性化需求。这一技术方案创新对于推进遥感数据共享具有深远的意义。

## 2.2 关键技术分析

### (1) MFS

在遥感数据云服务平台中，需要对大量的遥感

图像文件进行存储，并且能让其他计算机很容易地访问。在目前主流的分布式文件系统中，支持 FUSE 的分布式文件系统主要有 ceph, glusterFS, mooseFS。从技术上看，ceph 基于 btrfs 文件系统，保证了高性能，而多台 master 的结构解决了单点依赖问题，从而实现了高可用性。但 ceph 从技术层面上还不太成熟，不适合用于生成环境。glusterFS 比较适合大型应用，但由于没有元数据服务器，在压力比较大的时候性能大打折扣。MooseFS 虽然存在单点故障，管理服务器 master 比较占内存，但其部署简单、具有高可靠性和高性能的优点，刚好满足遥感数据云存储的需求。

在对三种分布式文件系统的性能对比时，本文设计了在三种不同文件系统下，大小文件读写的测试，实验如下：

① 硬件环境：分别 8 台机器，每台机器内存 32G，24 核 CPU，599G 硬盘空间；

② 软件环境：操作系统 Linux Fedora20x64 3.14.4-200.fc20.x86\_64，文件系统版本分别为 mfs-

表1 实验配置信息

Table 1 The experimental configuration information

	IP	节点配置	客户端IP
MooseFS	192.168.50.149~156	mfsmaster: 149 moosefs_backup-server: 150 chunkserver: 151~156	149~152
Ceph	192.168.50.10~17	MDS: 10~12 OSD: 13~17	10~13
GlusterFS	192.168.50.141~148	RAID1两两一组	142~145

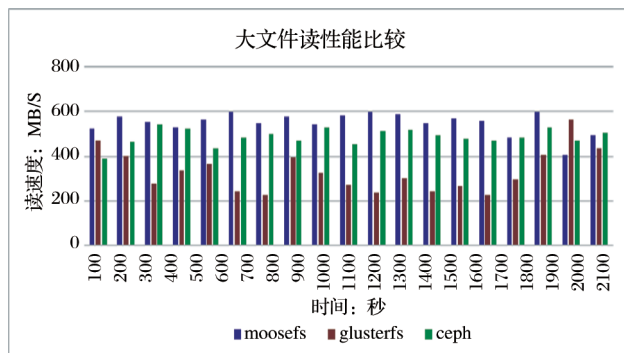


图8 大文件读性能测试结果

Fig. 8 Large file read performance test results

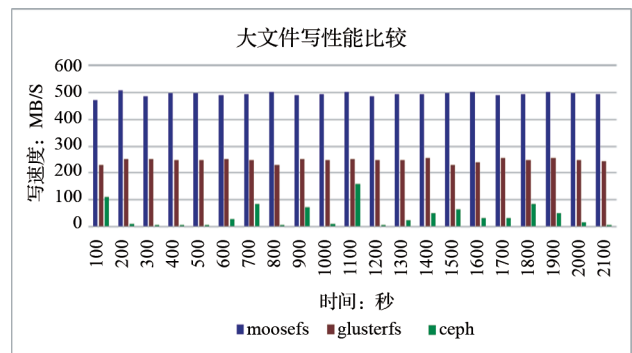


图9 大文件写性能测试结果

Fig. 9 Large file write performance test results

1.6.27-5、Ceph-0.80.1、glusterfs 3.5.0;

③ 配置信息: 三种文件系统均配置在 IP 网段 192.168.50.x, 具体配置信息见表 1;

④ 实验设计: 编写读写测试程序, 每个客户端分别启动 8 个进程运行测试程序, 分别向文件系统读写文件, 并记录所用时间, 文件大小取遥感影像数据文件大小平均值 100M;

⑤ 实验结果: 读写结果如图 8 和 9 所示。

从实验结果可以看出, Moosefs 在对遥感影像大小量级的文件读写速度均优于 Ceph 和 Glusterfs。在对三种主流的分布式文件系统进行对比后, MooseFS 成为目前遥感数据云服务平台中最合适的选择。

### (2) Docker

在遥感数据云服务平台中, 给用户提供了遥感数据的访问和计算资源是用户最主要的需求<sup>[9-10]</sup>。和传统的基于虚拟机的虚拟化技术相比, Docker 利用 LXC 实现了类似 VM 的功能, 从而利用更加节省的硬件资源提供给用户更多的计算资源。和 VM 方式不同, LXC 不是一套硬件虚拟化方法, 而是一个共享 Kernel 的操作系统级虚拟化方法。通过在执行时不重复加载内核, 且虚拟容器 (Container) 与宿主机 (Host) 之间共享内核来加快启动速度和减少内存消耗。根据实际测试结果, 基于 LXC 虚拟化的磁盘 I/O 和 CPU 性能接近裸机的性能, 大大优于虚拟机虚拟化技术。启动快且占用资源少, 对于遥感数据云服务平台这类轻量级的 PaaS, Docker 能够有效地构建隔离的标准化运行环境, 在有限的资源下提供计算资源给更多的用户。

### (3) IPython notebook

notebook 是 Web based IPython 的封装, 可以展现富文本, 使得整个工作可以以笔记的形式展现、存储, 对于交互式编程非常方便。用户基于 Web 可以在同一页面上混合 python 代码和文本, 让自己的代码块作为单独的程序运行。在遥感数据云服务平台中, 用户可以在 Web 上, 基于系统提供的遥感数据读写和通用算法接口编写代码, 对遥感图像文件进行访问计算并实时获取程序运行结果。随着系统通用算法的丰富和用户复杂算法的共享, 遥感数据处理和获

取的门槛将会慢慢降低。

### (4) ownCloud

ownCloud 是一个基于 Linux 的开源云计算项目, 允许用户以 Web 或 WebDAV 的方式访问云服务器, 进行各项管理工作。在遥感数据云服务平台中, 利用 ownCloud 可大大方便用户上传和下载文件, 包括用户运行的结果文件, 自定义上传文件等。

## 3 结论与展望

本文结合云计算技术和 Linux 虚拟化技术, 为遥感数据云服务提供了创新性的服务机制<sup>[11]</sup>。云计算技术将深刻改变遥感数据存储、处理技术和服务方式, 给遥感应用服务和产业化带来了重要机遇。Linux 虚拟化技术使得遥感数据服务满足大规模用户访问成为可能, 并大大降低了遥感数据的应用门槛。通过遥感数据云服务, 提供给用户通过 web 随时随地按需使用的平台, 解决了遥感应用中数据、技术、设备和成本的瓶颈, 为遥感数据应用的普及和产业化扫清了道路。另外, 该平台具有很好的移植性, 不仅可以运用于遥感数据领域, 也可以推广到其他大数据云服务平台等科研活动中。在本平台中, 完全代码级数据处理的交互方式仍然存在一定门槛, 如何研究基于互联网和图形窗口界面的服务方式需要进一步的设计和开发。最后, 本文作者还将对平台的安全性、稳定性做进一步研究和实验, 对平台数据高并发访问做进一步优化, 不断推进遥感数据云服务平台的发展。

## 参考文献

- [1] 武云龙, 王思勇, 李新楼. 2010. 基于云计算的遥感数据处理模型的设计与实现[J]. 电脑知识与技术, 6(14): 3646-3648.
- [2] Yang C, Goodchild M, Huang Q, Nebert D, Raskin R, Xu Y, Bambacus M and Fay D. 2011. Spatial cloud computing: how can geospatial sciences use and help to shape cloud computing. International Journal of Digital



- Earth, 4(4): 3 05–3 29 [DOI:10.1080/17538947.2011.587547].
- [3] 王昊鹏, 刘旺盛. 2008. 虚拟化技术在云计算中的应用初探[J]. 电脑知识与技术, 3(7): 1554-1565.
- [4] 董向军, 张恩刚, 张沛, 等. 桌面虚拟化技术研究[J]. 中国信息界, 2010, 140(4): 50~52.
- [5] 金海. 计算机虚拟化-原理及应用[M]. 北京: 清华大学出版社, 2008.
- [6] 冯登国, 张敏, 张妍, 徐震. 2011. 云计算安全研究[J]. 软件学报, 22(1): 77–83 [DOI: 10.3724/SP.J.1001.2011.03958].
- [7] 杨铁利, 许惠平. 2006. 网格技术在地理信息服务的应用研究[J]. 微电子学与计算机, 23(10): 141-149.
- [8] 邓书斌. 遥感信息一体化技术[R]. 中国区域用户大会 2010ESRI, 2010.
- [9] Golpayegani N and Halem M. 2009. Cloud computing for satellite data processing on high end compute clusters // Proceedings of the 2009 IEEE International Conference on Cloud Computing. USA: IEEE Computer Society: 88–92 [DOI: 10.1109/CLOUD.2009.71].
- [10] KIBEL S, WATANABE S, KUNISHIMA K, et al. PaaS on IaaS[C]. 2013 IEEE 27th International Conference on Advanced Information Networking and Applications, 2013.
- [11] Ren F H and Wang J N. 2012. Turning remote sensing to cloud services: Technical research and experiment. Journal of Remote Sensing, 16(6): 1331-1346.
- 收稿日期：2015 年 1 月 15 日
- 张品**：中国科学院计算机网络信息中心，硕士研究生，主要研究方向为分布式文件系统，大数据存储与处理。  
Email: [zp@cnic.cn](mailto:zp@cnic.cn)
- 张海明**：中国科学院计算机网络信息中心，博士，助理研究员，主要研究方向为 GNU/Linux 操作系统、分布式文件系统及大数据存储与处理。  
Email: [haim@cnic.cn](mailto:haim@cnic.cn)
- 黎建辉**：中国科学院计算机网络信息中心，博士，正研级高级工程师，博士生导师，主要研究方向为海量数据管理、处理与挖掘分析的理论、方法与关键技术。  
Email: [lijh@cnic.cn](mailto:lijh@cnic.cn)