



基于 MapReduce 的交互可视化平台^{*}

王加亮,秦 勃,刘健健,刘 妮

(中国海洋大学信息科学与工程学院 青岛 266100)

摘要

针对长时间序列的海洋环境信息交互可视化处理问题,本文提出一种基于云环境的交互可视化平台架构。可视化平台以 Hadoop 为基础,将 GPU、MPI 并行计算引入 MapReduce 处理机制,实现海洋环境空间数据场大规模数据集的检索、抽取、插值计算、特征可视化分析的并行处理,达到海量数据的远程交互可视化处理目的。实验结果表明,本文提出的可视化平台架构可以有效地将 MPI、GPU 并行计算与 MapReduce 处理机制相结合,以满足海洋环境信息交互可视化的需求。

关键词 海洋环境信息;云计算;GPU;可视化

1 引言

海洋环境信息主要包括海面高度、潮流、海流、海浪、温度、盐度、密度、气温、气压、湿度等。由于海洋环境的复杂性,各种指标非常多,空间和时间跨度非常大,其监测信息具有海量性的特点;并且海洋环境信息的观测手段多种多样,如浮标、台站、CODAS、CTD、ADCP、观测船、遥感、卫星等,导致数据的格式和精度多种多样,甚至以不同的数据形式存在,如图像、声音、文本等。由于这些信息基本以数字的形式呈现,很难作为海洋研究、开发利用和管理有效的分析手段,因此采用可视化处理技术对海洋环境信息进行分析、处理具有非常重要的意义。

鉴于海洋环境信息的海量性、多源性和数据形式多样性,海洋环境信息可视化分析处理需要采用大规模的并行计算模式以提高可视化分析、处理的效率,支持海洋环境信息的交互可视化需求。鉴于 Hadoop 云计算平台在海量数据检索、抽取等处理方面具有诸多优势,本文在已实现的

海洋环境信息并行可视化分析处理的基础上,将 GPU (graphic processing unit,图形处理器)^[1]、MPI(message passing interface)并行计算引入 MapReduce 处理机制,实现海洋环境空间数据场大规模数据集的检索、抽取、插值计算、特征可视化分析的并行处理,构建一个基于云环境的海洋环境信息远程交互可视化平台体系架构解决方案,达到海洋环境信息的温盐密度场、流场等的远程交互可视化的目的。

2 国内外研究现状

近年来,各大 IT 公司纷纷推出了自己的云计算基础设施和云计算服务,如谷歌的 Google Apps、微软的 Windows Azure、IBM 的 Blue Cloud 等,这大大促进了云计算技术的发展,使云计算技术在海量数据处理方面获得了广泛的应用。同时,为了解决流式处理的问题,各大 IT 公司也不断地对既有平台进行改造或者直接推出新的实时云计算平台解决方案,如 Facebook 的 Facebook Messages 就是在 Hadoop 基础上改造成的更有效的实时系统,而 Yahoo 则是推出新的 S4 开源流计算平台。我国云计算服务尚处于起步阶段,但是发展迅速,淘宝、360 安全卫士等都

^{*} 海洋公益性行业科研专项经费项目“海洋环境信息云计算与服务体系框架应用研究”(No.201105033)

在大规模使用和发展 Hadoop 云计算平台。但是目前将云计算理论应用于海洋环境信息远程交互可视化的工作还很少。

3 海洋环境信息交互可视化体系架构设计

一般情况下,云计算体系架构^[2]采用3层栈结构,分别为应用服务层(SaaS)、平台服务层(PaaS)、基础设施服务层(IaaS),由不同的层次提供不同级别的服务。海洋环境信息交互可视化云计算体系架构(见图1)也分为3层:应用层(包括用户接口层、应用模型层)、平台层(包括云服务层、云平台)、资源层(包括虚拟资源层、物理层)。

(1)应用层

用户接口层主要接收用户的交互信息,形成任务描述文件,为用户显示可视化任务执行结果等。应用模型层主要是各种定制的可视化模型,其响应用户接口层的各种任务请求,利用云平台提供的各种云服务完成具体的可视化任务,并为海洋环境信息采集用户提供数据上传服务。

(2)平台层

云服务层主要完成对各种可视化定制模块处理任务

的响应,对应海洋环境信息可视化的数据预处理、数据的映射、可视化图像的绘制3个步骤,提供各个可视化服务分割模块的具体处理服务。云平台层主要提供具体的存储服务 and 计算服务,并对 Hadoop 的计算服务进行了一定的改进,引入了 GPU 并行计算、MPI 并行计算、CPU 并行计算等多种类型、多种层次、多种粒度的并行模式,使其更适应海洋环境信息数据交互可视化任务的需求,在第4节将做更具体的陈述。

(3)资源层

资源层主要为平台层提供计算存储服务。其中,物理层主要由磁盘阵列组成的数据中心以及配有大容量硬盘的 CPU 计算服务器和 GPU 计算服务器组成;虚拟层通过虚拟化技术,将来自不同节点、不同结构的物理资源进行整合,形成大型资源池供平台层使用。

4 MapReduce 处理机制优化

4.1 并行计算模式

并行计算模式依照数据和程序的结合过程分为3种

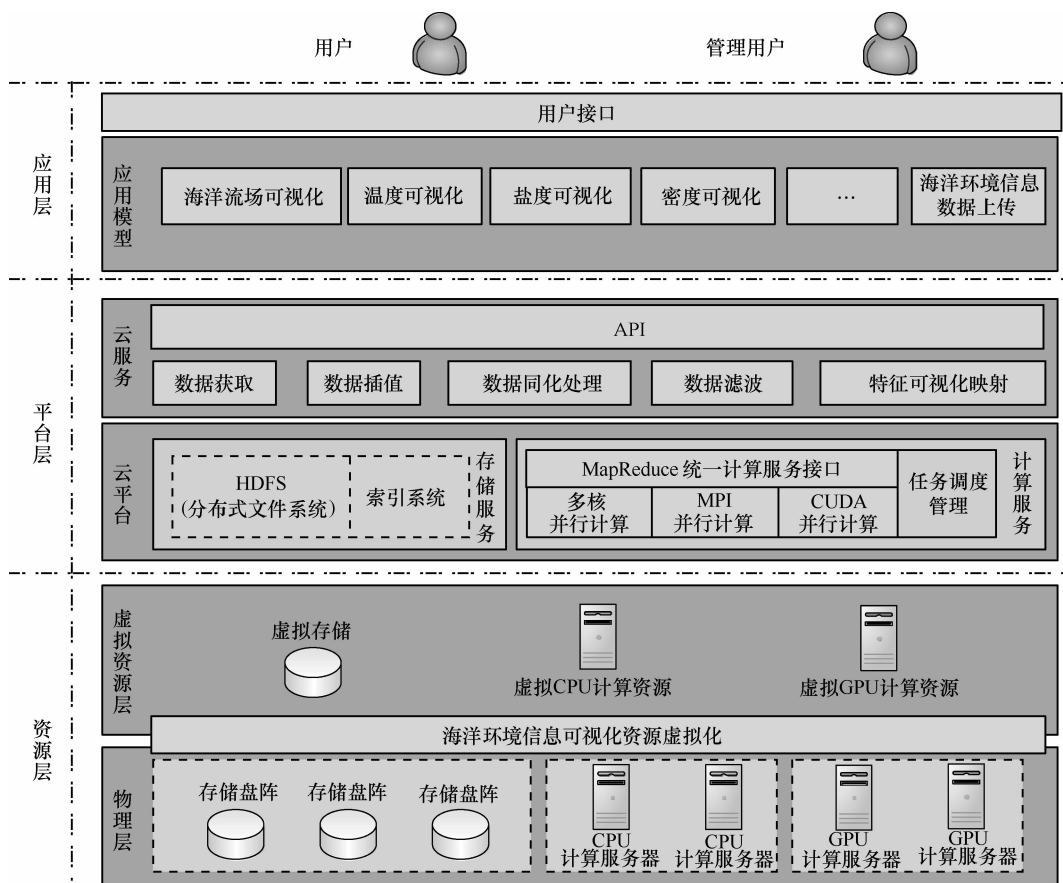


图1 海洋环境信息可视化云计算体系架构



方式:一是将数据移到程序所在的位置,如传统的 MPI;二是将程序移动到数据所在的位置,如 Hadoop 云计算平台的 MapReduce;三是将程序和数据共同移到第三方位置,如 GPU 并行计算。

MPI 是高性能计算(high performance computing, HPC)和网格计算(grid computing)采用的主要方式。其将作业分散到集群的各台机器上,这些机器访问由存储区域网络(SAN)组织的共享文件系统,适用于计算密集型的作业。如果节点需要访问更大量的数据(几百个 GB 的数据),那么很多计算节点会由于网络带宽的瓶颈问题而空闲下来等待数据。

GPU 将大量的晶体管用作 ALU 计算单元,具有强劲的计算能力、高性能/价格比和高性能/能耗比,再加上其 3 倍于摩尔定律的发展速度,在当今追求绿色高性能计算的时代,GPU 的计算优势受到越来越多的关注。除专业图形应用外,GPU 已大量地用于通用计算问题,并形成了 GPU 通用计算研究领域,即 GPGPU (general-purpose computing on graphics processing units),又称 GP2U。GPU 相对独立的并行计算架构,非常适合处理海洋环境信息可视化这种兼具数据密集和计算密集特点的任务处理。

Hadoop 是一个能够对大量数据进行分布式处理的软件框架,它是 Google 的 MapReduce 编程模型和框架的开

源实现,其可以在大量廉价的硬件设备组成的集群上运行应用程序。它为应用程序提供了一组稳定可靠的接口,是一个具有高可靠性和扩展性的分布式系统,在海量数据处理方面性能很强,并在地震模拟、互联网信息挖掘等方面得到了广泛的应用。

综观上述并行计算模式及海洋环境信息海量数据的特点,可以看出交互可视化处理无论是 GPU 还是 MPI,并行计算的通信时间消耗都会急剧地增加,而且单 GPU 的运算能力也是有限的;而 Hadoop 默认模式被设计成适合海量数据处理,侧重于最大化吞吐量和效率,最初被设计成适合处理离线且 I/O 方式是从磁盘上顺序读写大量数据的批处理式工作,对于随机读写和实时在线需求的支持多是采用将 MapReduce 的批处理结果上传到 MYSQL 等方式来实现近乎实时的低时延响应。

为了建立一种更高效的处理海量环境信息的交互可视化体系架构模式,4.2 节将结合 MapReduce 移动运算到数据所在节点,减少海量数据通信和 GPU 运算较高加速比的优点,提出一种基于 MapReduce 的多层次并行处理机制,对 MapReduce 的处理机制进行优化。

4.2 基于 MapReduce 的多层次并行处理机制

基于 MapReduce 的多层次并行处理机制如图 2 所示,该机制^[3,4]中融合了 GPU 和多核 CPU 多种计算资源,首先

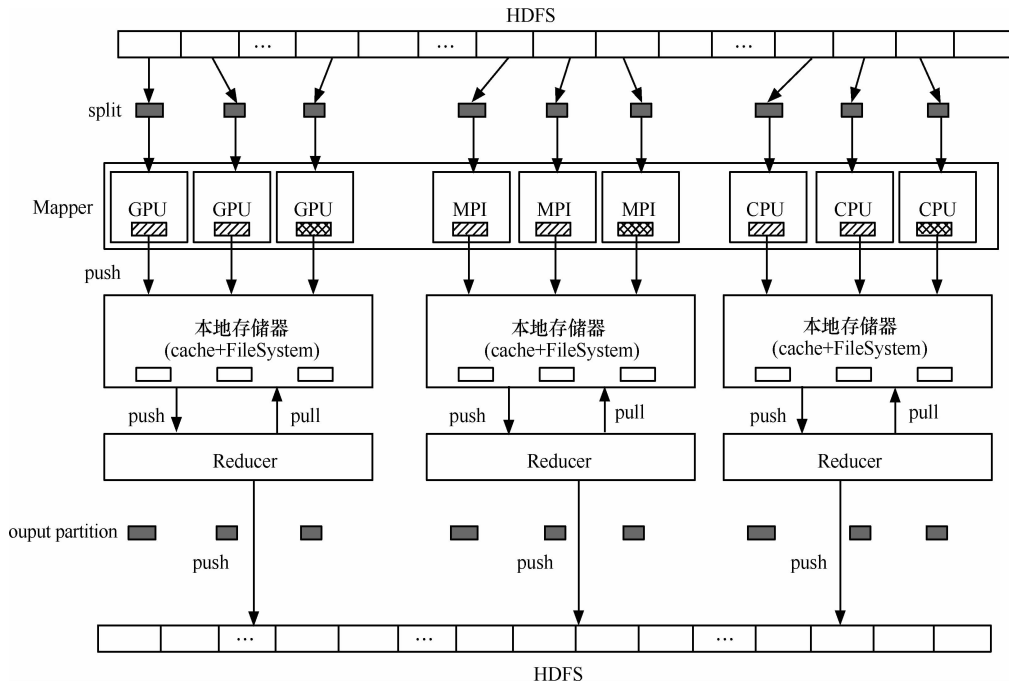


图 2 基于 MapReduce 的多层次并行处理机制

通过 MapReduce 机制对任务进行分发,实现第一层次多节点的并行,然后在每个节点上根据节点的配置情况采用 GPU 等并行方式实现第二层次的并行,这种多层次并行处理机制可以有效发挥目前广泛存在的混合异构集群的计算能力。

(1) MapReduce 的多层次并行体系

针对海洋环境信息可视化数据密集的特点,由于 Hadoop 默认机制是数据本地化(data-local)的,也就是任务运行在输入分片所在的节点上,或者机架本地化(rack-local):即任务和输入分片在同一个机架上,但不在同一个节点上。对于数据密集型的运算,这可以大大减少节点间的通信负载,因此应用 Hadoop 能较好地解决海洋环境信息数据海量的问题。

同时,针对海洋环境信息可视化计算密集的特点,鉴于 GPU 非常适合高计算强度的应用,因此将 GPU 应用到 MapReduce 过程中^[6],即在 Hadoop 的并行基础上增加一个并行层次,实现多层次并行。另外,实验结果也显示:高计算强度是发挥 GPU 高浮点计算性能的前提条件,否则由于 CPU 与 GPU 之间的数据传输及同步开销过大,不能发挥 GPU 的优势。具体实现时,应优化算法减少数据传输的次数以及数据量和同步的开销。但是也应该看到,如果数据粒度过小,那么计算强度也会成比例地下降,因此应该通过实验,寻找一个最优的数据粒度。将 GPU 应用到 MapReduce 过程中,采用大粒度的 CPU 并行预处理数据,大规模计算采用 GPU 进行第二层次的并行,可以提高 Hadoop 集群的运算效率和吞吐率。

MPI 也比较适合计算密集型的作业,通过对 MPI 进行改进减少通信开销,MPI 能够应用于数据密集型计算领域,并能显著加速 Hadoop 和 MapReduce 应用程序^[7]。

(2) 基于管道的数据传送

Hadoop 系统^[7]默认将每个 Map 和 Reduce 阶段的全部输出在下阶段被消耗前就被实体化到本地稳定存储器或者产生输出,虽然批量实体化可以简化 Hadoop 的容错机制,其对于大型部署十分关键,但是这会导致运算节点的怠工,因此可以在 MapReduce 之间增加一个管道,变 Hadoop 默认的“推”(即在 Map 阶段运算完毕产生中间结果之后才开始 Reduce 阶段)为“推”和“拉”(即在 Map 阶段没有完全计算完毕产生部分中间结果的时候就开始 Reduce 阶段的运算)并举,这种方式可以增加并行机会,提高利用率,从而缩短工作完成时间,最终减少响应时间。

5 海洋环境信息可视化云计算平台的构建与实现

5.1 海洋环境信息可视化云计算平台的构建

海洋环境信息可视化云计算实验平台网络及物理结构如图 3 所示。实验平台的硬件配置主要是 Hadoop 集群中的 12 台机器,每台机器配有 Intel Core i5-23020 CPU,4 GB 内存,其中 8 台机器配有支持 CUDA 运算的 Nvidia GTX 460 显卡,另外 4 台机器的显卡不支持 CUDA 运算,其中一台机器作为 Hadoop 的主节点,其他为从节点,并通过吉比特网络连接曙光 TC 4000A 集群。在 Hadoop 集群之外配有门户服务器、应用服务器和管理服务器。

软件配置方面主要是为每台机器安装配置了 Hadoop 0.20.2、MPICH2、Nvidia 的 CUDA Toolkit 4.1。另外为每台机器配置了 Jcud 0.4.1。

5.2 海洋环境信息可视化处理流程及其实现

海洋环境信息可视化处理流程如图 4 所示。用户通过浏览器访问云计算门户 Web 服务器,然后再由 Web 服务器调用 Hadoop 集群进行大规模并行计算,Hadoop 集群的执行结果经 Web 服务器最终在客户端的浏览器中以网页的形式呈现。

(1) 可视化计算过程调用及结果显示的实现

Web 服务器对 MapReduce 计算的调用可以用后端执行 Shell Script 的方式实现。在可视化结果的展示方面,使用基于 HTML5 的 WebGL 技术,WebGL 是一种 3D 绘图标准,这种绘图标准允许把 JavaScript 和 OpenGL ES 2.0 结合在一起,通过增加 OpenGL ES 2.0 的一个 JavaScript 绑定,WebGL 可以为 HTML5 Canvas 提供硬件 3D 加速渲染,其硬件加速功能可以将可视化数据更快地呈现给用户,并更快地响应用户的交互请求。另外,虽然从云计算理论上讲,希望能够构造一种弱客户端的应用,即尽量将存储和计算资源都集中于庞大的“云”中心,但是推送运算到数据并不能完全适合所有的情况,客户端的运算并不能完全由云计算所替代,海量数据分析的应用程序能够从“服务器+客户端”的体系结构中受益^[8]。因此为了加强用户的交互性,在用户软硬件环境支持的情况下将一部分“小数据”的任务在本地客户端利用 WebGL 的硬件加速功能来完成,实验证明本地化处理是一种简单的、有利于减少时延和公共资源数据加载的有效方法。

(2) 多层次并行处理机制实现

实验中使用的开发语言为 Java 编程语言。由于 Nvidia

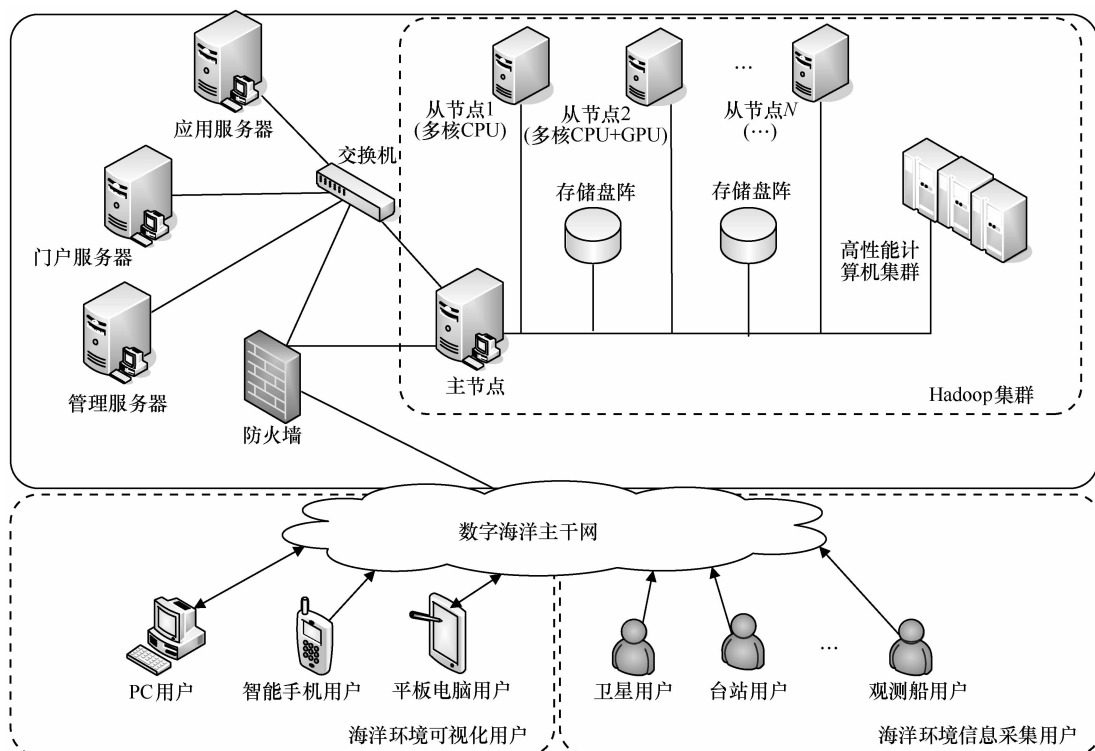


图3 海洋环境信息可视化云计算平台网络及物理结构

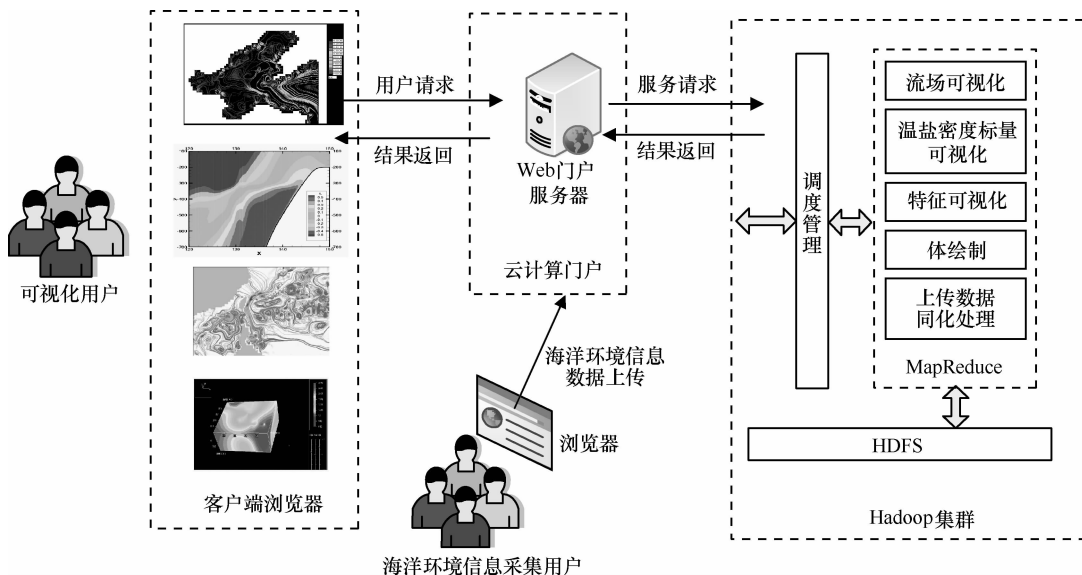


图4 海洋环境信息可视化数据处理流程

的 CUDA 编程框架目前只支持 C 和 C++ 语言, 为了实现 Java 对 GPU 的调用引入 JCuda, JCuda 是对 CUDA runtime 和 driver API 的一种 Java 封装, 其实现了 Java 对 GPU 的调用。具体的调用方式一种是使用 jcudaUtils-0.0.4.jar 中的相关接口, 将 CUDA 代码封装为字符串, 这种方式的优点是实现对 CUDA 代码的自动编译和向 TaskTracker 的自动

分发, 缺点是在每次执行过程中都会对 CUDA 代码进行编译, 增加了额外开销; 另一种方式是使用 JNI (Java native interface) 调用编译好的 CUDA 代码, 这种方式的优点是 CUDA 代码只需要编译一次, 但是需要手动将编译好的 CUDA 二进制代码放到每台 TaskTracker 上, 而且使用 JNI 也可能会丧失代码的平台移植性。

另外,GPU 的初始化是需要时间的,因此为了有效发挥 GPU 的优势,在 Map 中加入 GPU 运算时粒度不应该太小,但是由于 Hadoop 本身是通过启动 TaskTracker 的一个 Java 线程来执行任务的,如果大粒度的任务数据量过大的话,会出现 JVM 内存溢出的错误,为此需要加大 Java JVM 的 Heap space 的大小。

6 结束语

本文针对海洋环境信息长时间序列海量数据的特点,提出了一种基于 Hadoop 云计算平台的海洋环境信息远程交互可视化的体系架构,并针对 MapReduce 进行了优化,采用多级多粒度的并行运算体系架构,提高了海洋环境信息可视化的响应效率。但是由于 Hadoop 本身的设计目标是应用于大规模海量数据的批处理式工作,因此对 Hadoop 相关配置及多级并行运算算法进行优化,提高其运算处理效率及响应速度将是未来工作的重点。

参考文献

- 1 Buck I. GPU computing: programming a massively parallel processor. International Symposium on Code Generation and Optimization(CGO'07),California,2007:17~23
- 2 曾诚,李兵,何克清.云计算的栈模型研究.微电子学与计算机,2009,26(8):22~27
- 3 Polo J, Carrera D, Becerra Y, *et al.* Performance of accelerated MapReduce workloads in heterogeneous clusters. Proceedings of 39th International Conference on Parallel Processing, San Diego, 2010:653~662
- 4 卢风顺,宋君强,银福康等. CPU/GPU 协同并行计算研究综述.计算机科学,2011,38(3):5~9
- 5 Huy T Vo, Broson J, Summa B, *et al.* 2011 IEEE Symposium, RI, 2011:81~89
- 6 Condie T, Conway N, Alvaro P, *et al.* MapReduce OnLine, UCB/EECS-2009-136. Berkeley: Electrical Engineering and Computer Sciences University of California, 2009
- 7 Lu Xiaoyi, Wang Bing, Zha Li, *et al.* Can MPI benefit Hadoop and MapReduce applications. Proceedings of 2011 International Conference on Parallel Processing Workshops, Taipei, China, 2011:371~379
- 8 Crochow K, Howe B, Stoermer M, *et al.* Client+Cloud:evaluating seamless architectures for visual data analytics in the ocean sciences. Proceedings of 22nd International Conference on Scientific and Statistical Database Management, Berlin, 2010: 114~131

Interactive Visualization Platform Based on MapReduce

Wang Jialiang, Qin Bo, Liu Jianjian, Liu Ni

(College of Information Science and Engineering, Ocean University of China, Qingdao 266100, China)

Abstract For dealing with the problem of ocean data's interactive visualization, this paper proposes an interactive visualization platform architecture based on cloud computing. It inserts GPU, MPI parallel computing into MapReduce mechanism of Hadoop to realize the parallel processing of large-scale ocean environment data sets, such as data retrieval, data extraction, data interpolation, analysis of characteristics' visualization, so that the massive data's visualization can be processed in the remote interactive way. The experiment results show that the visualization platform architecture can combine MPI, GPU with MapReduce effectively to meet the needs of ocean data's interactive visualization.

Key words ocean environment information, cloud computing, GPU, visualization

(收稿日期:2012-07-20)