

海洋环境信息云计算应用研究

华彦宁¹, 周雪¹, 陈艳云¹, 卢峰², 陈刚¹, 张镭¹

1. 国家海洋信息中心, 天津 300171

2. 中电科海洋信息技术研究院有限公司, 北京 100041

摘要 在总结国内外云计算发展及应用现状的基础上, 结合海洋环境信息系统的特点, 对海洋环境信息云计算的体系架构及相关技术进行了研究, 设计了海洋环境云平台; 针对海洋环境信息联机分析处理、海洋环境数据挖掘服务、海洋遥感影像处理与关键因子反演、乘潮水位预报系统等典型应用开展了示范服务的实践。

关键词 海洋环境信息; 云计算; 云平台; 虚拟化

云计算 (cloud computing) 是随着互联网的高速发展, 基于分布式计算 (distributed computing)、并行计算 (parallel computing)、网络存储 (network storage technologies)、虚拟化 (virtualization)、负载均衡 (load balance)、热备份冗余 (high available) 等传统计算机和网络技术, 逐渐发展融合的产物^[1-3]。目前对于云计算并没有一个准确的定义, 较为广泛接受的是美国国家标准与技术研究院 (National Institute of Standards and Technology, NIST) 对云计算的定义。云计算是一种按使用量付费的模式, 这种模式提供可用的、便捷的、按需的网络访问, 进入可配置的计算资源共享池 (资源包括网络、服务器、存储、应用软件、服务), 这些资源能够被快速提供, 只需投入很少的管理工作, 或服务供应商进行很少的交互^[4-5]。

在当前海洋环境信息领域的发展中, 资源占用多、运行成本高, 传统模式难以适应业务部署的快速要求, 缺乏统一部署计算资源的规划; 业务系统稳定性和可

靠性低、系统维护难度大等问题逐渐凸显。云计算技术则能够可以有效地解决上述海洋信息系统建设中存在的问题, 并日益受到涉海机构部门的重视。本文在综合国内外云计算发展现状的基础上, 分析总结了适应海洋环境信息的云计算体系架构, 针对性地对资源虚拟化和云平台构建技术进行了研究, 结合海洋环境信息云平台示范系统的实践成果, 展望了未来云计算在海洋领域的发展趋势。

1 国内外云计算发展现状

1.1 国外云计算的发展

云计算的概念最早由 Google 首席执行官 Eric Schmidt 于 2006 年 8 月 9 日在 San Jose 召开的搜索引擎大会 (Search Engine Strategies Conference, SES) 上首次提出^[6]。受到政府机构、国际组织、公益服务部门、大型企业的极大关注, 并在产业界和商业界快速得到了成

收稿日期: 2018-06-22; 修回日期: 2018-07-02

作者简介: 华彦宁, 高级工程师, 研究方向为海洋信息技术, 电子信箱: yanminghua@hotmail.com; 张镭 (通信作者), 高级工程师, 研究方向为海洋信息技术, 电子信箱: zhanglei@nmdis.org.cn

引用格式: 华彦宁, 周雪, 陈艳云, 等. 海洋环境信息云计算的应用研究[J]. 科技导报, 2018, 36(14): 42-56; doi: 10.3981/j.issn.1000-7857.2018.14.007

功应用。虽然不同企业的云计算应用针对的用户需求不同,其实现的架构和技术各有特点,但都具有基于网络技术的发展、资源虚拟化、分布式系统和按需使用等特性。

1) Amazon的“弹性计算机云”体系架构与运行机制。

Amazon于2006年3月推出弹性计算云(elastic compute cloud, EC2)服务,早于“云计算”概念的提出,可以将其看作是云计算的雏形^[7]。弹性计算云在实体计算机群上构建了一个虚拟的集群环境,弹性计算云中的每一个实例代表一个运行中的虚拟机。用户对自己的虚拟机具有完整的访问权限,包括针对此虚拟机操作系统的管理员权限。虚拟机的收费也是根据虚拟机的能力进行费用计算,实际上,用户租用的是虚拟的计算能力。在用户具有充分灵活性的同时,也减轻了Amazon公司的管理负担。

2) Google的云计算体系架构与运行机制。

Google云计算主要由分布式文件系统(google file system, GFS)、大规模分布式数据库(Big Table)、程序设计模式(Map/Reduce)、分布式锁机制(Chubby)等几个相互独立又紧密结合的系统组成^[8-9]。该系统通过GFS实现存储,采用Big Table技术作为其数据库,实现结构化、半结构化数据存储;通过Map/Reduce来处理 and 产生大量数据集的相关实现,用于大规模数据集(>1 TB)的并行运算,使程序员不需要任何并行以及分布式系统的经验,就能够轻松地使用大型分布式系统的资源;Chubby技术用来进行云计算的互斥和同步,实现进程锁控制管理。

3) IBM的“蓝云”计划的体系架构与运行机制。

“蓝云(Blue Cloud)”计算平台由数据中心、IBM Tivoli 监控软件、IBM DB2 数据库、IBM Tivoli 部署管理软件、IBM WebSphere 应用服务器以及开源虚拟化软件和一些开源信息处理软件共同组成^[10]。该平台采用了Xen、PowerVM 虚拟技术和Hadoop 技术帮助客户构建云计算环境。其中,Hadoop是由Apache Software Foundation 公司基于Google公司Map/Reduce和GFS而开发的一个能够对大量数据进行分布式处理的软件框架。“蓝云”软件平台的主要特点体现在其成功结合了虚拟机、大数据分布式处理软件Hadoop和Xen虚拟平台,实现了高可靠、低成本的数据应用服务和计算能力的可扩展。

4) Yahoo云计算的体系架构与运行机制。

Yahoo云计算采用Hadoop技术构建在大型集群的廉价硬件设备上运行应用程序的框架^[11]。利用Hadoop框架,为应用程序提供了一组透明、可靠的服务和数据接口,将应用程序分割成若干工作单元,每个单元可以在集群任何节点上执行或重复执行;该框架提供一个分布式文件系统用于在各个计算节点上存储数据,并实现了对数据读写的高吞吐率;通过采用Map/Reduce和分布式文件系统使得Hadoop框架具有高容错性,能够自动处理失败节点。Yahoo云计算已经通过了600个节点的集群测试,并具有进一步的扩展性。

5) 美国国家海洋与大气管理局主导的海洋环境数据云服务。

通过与开放云联盟(Open Cloud Consortium, OCC)合作,美国国家海洋与大气管理局(National Oceanic and Atmospheric Administration, NOAA)利用云计算技术,加强其环境数据产品的服务,向公众提供更多地NOAA的环境数据,使用户更容易查询、访问数据,进行研究和分析^[12]。其搭建的开放科学数据云(open science data cloud, OSDC)平台,包括环境数据、存储空间、计算能力和基础分析工具等资源,并以通用接口的形式对用户提供服务。

1.2 国内云计算的发展

与国外相比,中国的云计算起步稍晚,但发展较快。2008年开始,通过与国外公司合作的方式引入了“云计算”的概念并迅速投入应用。

2008年5月10日,IBM公司(International Business Machines Corporation)在中国无锡太湖新城科教产业园建立的中国第一个云计算中心,并投入运营;2008年6月24日,IBM公司在北京IBM中国创新中心成立了第二个中国的云计算中心——IBM大中华区云计算中心;2008年11月28日,广东电子工业研究院在东莞松山湖投资2亿元人民币建立云计算平台;2008年12月30日,阿里巴巴集团旗下子公司阿里软件与江苏省南京市政府正式签订战略合作框架协议,并于2009年初在南京建立国内首个“电子商务云计算中心”^[13];2009年3月,世纪互联数据中心有限公司推出了CloudEx产品线,包括完整的互联网主机服务“CloudEx Computing Service”和基于在线存储虚拟化的“CloudEx Storage Service”,供个人及企业进行互联网云端备份的数据保全服务等系列互联网云计算服务^[14]。

2010年之后,国内云计算进入实质性发展的阶段,主要体现在5个方面。

1) 大型企业建立的公有云。

包含了如IBM、EMC、Intel等为代表的跨国设备制造商转型为云计算解决方案提供商,也包含以新浪、腾讯、阿里巴巴、世纪互联等为代表的国内互联网企业,以较低成本和较高性能为公众提供海量信息存储和计算的解决方案。

2) 由地方政府主导建设的云计算中心。

以上海、北京、天津、无锡、东营等为代表的地方政府主导建设了一些云计算中心,建立政府政务云及面向中小企业的公有云,主要为大量的中小企业提供政务、资源、技术等多方位的服务。

3) 大型企事业单位的私有云建设。

以中国移动通信集团公司、中国电信集团公司为代表的传统电信运营商,为运营支撑系统搭建私有云,整合内部资源,节能降耗,实现利旧和转型。

4) 云安全领域的研究。

国内在云安全方面的研究处于世界前列。通过大量分布的客户端对网络中软件行为的异常监测,以获取互联网中木马、恶意程序的最新信息,搜集到服务端进行自动分析和处理,再把研发的病毒和木马的解决方案分发到每一个客户端。目前,国内杀毒软件的大公司都推出了云安全解决方案。

5) 云计算相关的前沿技术研究。

目前国内一些新兴的高科技公司,将云计算与大数据、物联网、智能芯片等前沿技术相结合,开展了一系列的下一代技术研究。2018年初,智能芯片领域的寒武纪公司已经推出新一代机器学习处理器云端智能芯片MLU100,针对智能手机的云端和终端两大应用场景提出了新一代云计算方案。

由于中国海洋领域信息化建设相对滞后,云计算技术尚未得到充分的应用,但已经得到了相当的重视,并开展了相应的应用研究。目前,云计算的应用大多数以私有云的形式,在一个单位内部,以提高硬件设备使用率、集中对外提供数据服务为目的开展建设。例如国家海洋环境监测中心已经在开展海洋环境监测云服务中心的建设,集成已有的海洋环境监测系统与业务,并对用户开展监测数据服务^[15]。

1.3 云计算的基本结构

从目前云计算的研究和应用来看,云计算主要分

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)3种类型^[16-17]。不同的厂家根据业务需求、服务模式的不同,提供了更加细化的解决方案^[18]。从表面上看,云计算的技术千差万别,但无论是哪种类型,其区别主要在于提供的服务和提供服务的方式有所不同,其基本结构和主要特征是类似的。传统的云计算体系结构概括了不同解决方案的主要特征,每一种方案或许只实现了其中部分功能,或许额外提供其他相对次要的功能。

云计算由4个部分组成(图1):物理资源、资源池、管理中间件和面向服务的体系结构(service-oriented architecture, SOA)构建^[19]。物理资源包括计算机、存储器、网络设施、数据库和软件等;资源池是将大量相同类型的资源构成同构或接近同构的资源池,如计算资源池、数据资源池等,构建资源池更多是物理资源的集成和管理工作;管理中间件是负责对云计算的资源进行管理,并对众多应用任务进行调度,使资源能够高效、安全地为应用提供服务;SOA构建将云计算能力封装成标准的Web Services服务,并纳入到SOA体系进行管理和使用,包括服务注册、查找、访问和构建服务 workflow等。管理中间件和资源池层是云计算技术的最关键部分,SOA构建层的功能更多依靠外部设施提供。

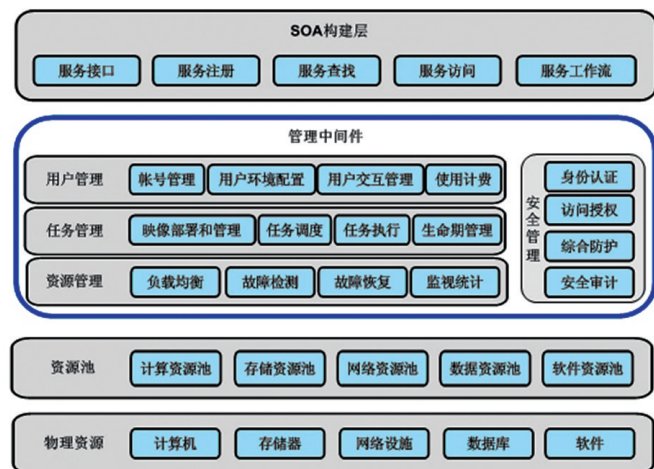


图1 云计算技术体系结构
Fig. 1 Cloud computing architecture

1.4 云计算的关键技术

从云计算基本结构可以看出,云计算需要将大量的各类资源进行有效整合和管理,为用户提供高效、标准的服务。结合当前主流的云计算应用实例,云计算

的关键技术包括虚拟化技术、分布式存储技术和分布式计算技术等。

1) 虚拟化技术。

根据资源的类型不同,虚拟化技术主要包括服务器虚拟化、存储虚拟化、应用虚拟化和桌面虚拟化^[20]。

服务器虚拟化是IaaS的核心技术之一,将一个物理服务器虚拟成若干个独立的虚拟服务器使用,充分发挥服务器的硬件性能。通过将CPU、内存、设备等物理资源转化为统一管理的逻辑资源,为每一个虚拟服务器提供支持其运行的抽象资源。服务器虚拟化技术能够将物理服务器的利用率从5%~20%提高到85%~90%^[21],大幅度减少了应用系统所需服务器的总体拥有成本。有利于在需求变化时,快速改变系统配置,动态分配系统所需的计算资源。

存储虚拟化是云存储系统中非常重要的一部分,实现对大量存储设备中的统一管理。在云存储中,用户并不关心数据具体存储位置,而是关心如何使用存储的数据,数据将如何呈现到系统终端,数据存储是否安全等。相比单纯增加物理存储设备,提高存储设备的利用率是降低成本的同时实现扩展容量的最佳方法。存储虚拟化的关键是解决不同存储设备的物理差异性,以及不同数据的统一存储及管理;同时还使存储空间具有可伸缩性,能够实现用户对存储容量的动态扩展和动态分配^[22]。

应用虚拟化技术也是SaaS的核心技术。通过一个虚拟层,可以提供所有与应用有关的注册表信息、配置文件等,使所有应用都可以运行,同时应用被重新定位到一个虚拟的位置,与只跟本身有关的运行环境打包,形成一个单一文件,可以在不同的环境下运行。这样的云计算平台能够在同一环境下支持不兼容的应用同时运行,避免不同应用间产生的动态链接库冲突。另外,采用应用虚拟化技术,可以在数据中心上集中管理打包的虚拟应用,当需要改变应用部署时,无需重新安装应用程序,只要通过在数据中心下载即可实现。

远程桌面是桌面虚拟化的原型,但是它只允许用户通过网络访问安装在另一个实体机的操作系统,不适合大规模应用。桌面虚拟化技术将操作系统、应用和用户配置文件等信息在数据中心内打包,并作为单一镜像动态地为每位用户装配虚拟桌面,用户可以随时随地通过任何设备进行智能身份验证,访问他们的个人计算机(PC)桌面^[23]。采用桌面虚拟化技术,可以

将用户的桌面操作环境与其使用的终端设备相分离,在数据中心的服务器中存放每个用户的完整桌面环境,满足用户在任何地点,通过任何设备访问和操作桌面环境的需求。

2) 分布式存储技术。

云计算系统由大量服务器组成,同时为用户提供服务,为保证高可用、高可靠和经济性,云计算采用分布式存储的方式来存储数据,用冗余存储的方式保证数据的可靠性。目前云计算系统中常以Google的文件系统(google file system, GFS)或Hadoop项目下的文件系统(hadoop file system, HDFS)作为分布式文件系统,其中HDFS是GFS的一个开源实现。由于云计算通常需要对海量的数据存储、读取后进行大量分析,数据的读写操作频率远远大于数据的更新频率,一般的数据库管理系统难以达到性能要求。因此,云计算普遍采用列存储的数据库管理机制,将数据表按列划分后进行分布式的存储。目前云计算系统中常以Google的BigTable或Hadoop项目下的HBase(Hadoop DataBase)作为分布式数据库^[24]。云计算环境下的数据库应具有高可靠性、高性能、面向列、可伸缩的特点,适应在廉价PC上搭建起大规模结构化的存储集群。

由于云计算分布式存储的特点,与一般的数据库技术相比,在云计算环境下,对数据进行分片化组织、优化索引与并行查询技术、增强节点间负载均衡的能力等就显得尤为重要。

3) 分布式计算技术。

在云计算环境下通常采用分布式存储,数据集一般分散在各个存储节点上,且任何大型的计算处理业务,也会通过服务器虚拟化,由诸多节点共同完成,这就需要通过分布式计算技术来解决。目前最为常见的分布式计算技术是Google提出的MapReduce模型^[25],用于大规模数据集的并行运算(图2)。

MapReduce采用“分而治之”的思想,把对大规模数据集的操作分发给一个主节点管理下的各分节点共同完成,然后通过整合各分节点的中间结果,得到最终的结果。这样的处理过程被MapReduce高度地抽象为Map和Reduce这两个函数。Map负责把任务分解成多个任务,Reduce负责把分解后多任务处理的结果汇总起来。在并行编程中的其他各种复杂问题,如分布式存储、工作调度、负载均衡、容错处理、网络通信等,均由MapReduce框架负责处理。

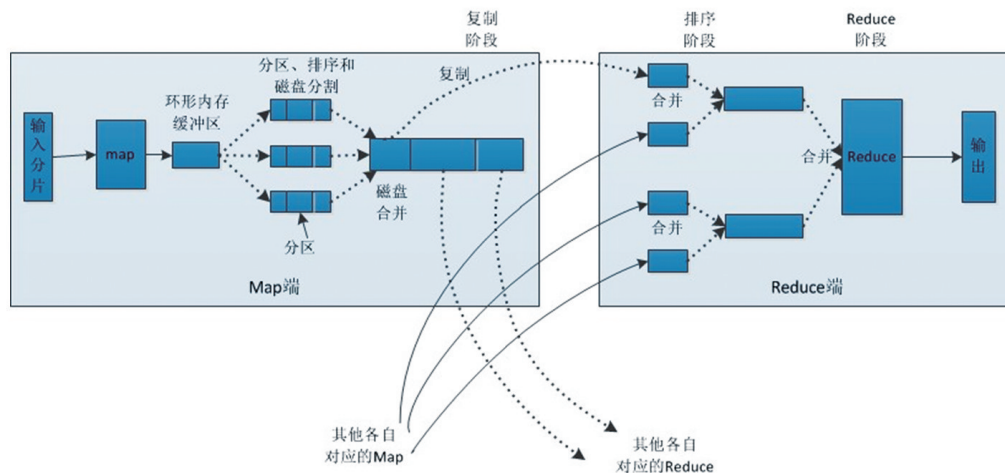


图2 MapReduce 模型
Fig. 2 MapReduce model

Map 和 Reduce 函数的具体功能由用户按照需求自行实现。在 Map 阶段,待处理的数据将被分解为固定大小的 split,为每个 split 创建一个 Mapper,用于执行用户定义的操作。在 Reduce 阶段,将不同 Mapper 接收的数据结果合并及排序,调用用户定义的 Reduce 操作,并得到最终结果。

2 海洋环境信息云计算的体系架构研究

2.1 海洋环境信息及海洋环境信息系统的特点

海洋环境信息的特点主要有时空性、来源多样、种类繁多及边界模糊等^[26]。

1) 时空性。海洋环境信息具有陆地环境信息的复杂性和流体信息的独特性的双重特点,是一种时空数据,具有变化尺度多样性的特点。因此,海洋环境信息既包含瞬时变化的信息(如海啸、火山爆发等),也有年度变化的信息(如潮流等),也包含有万年变化的信息(如地质构造等)。

2) 来源多样。海洋环境信息获取困难、成本高昂,不同国家、地区、机构有各自的信息获取计划,且信息采集设备种类繁多。因此,海洋环境信息的来源多样、标准不一,包括测量方式、坐标系、精度、观测尺度等,其数据多以数据集的形式存在,满足不同领域、不同业务的应用需求。

3) 种类繁多。海洋环境信息包括遥感影像信息、地形信息、矿产资源信息、水产资源信息、水文气象信息和海洋大地物理信息等诸种类,不同种类的数据

又通常具有较强的相关性,造成海洋环境信息交叉应用较多。

4) 边界模糊。海洋中地理区域的边界界限(如领海界线、大陆架、海洋经济开发区、禁止捕捞区、养殖区等)无法像陆地区界线那样精确和清晰,其呈现出一定的模糊性。而且不同数据,即使是相同的观测要素,也存在大量的边界重叠和空白区域,对信息的综合应用造成一定的困难。

与一般的信息化系统相比,海洋环境信息系统一方面依赖于海量的海洋环境信息,其数据组织、存储方式等受限于海洋环境信息的特点;另一方面,由于海洋环境信息系统的应用范围较为专业化,因此其侧重点与关键技术也有别于一般的信息系统。

因此,海洋环境信息系统的主要特点有:与地理信息系统深度融合;对数据存储、处理能力的要求较高;交叉应用日益增多。

1) 与地理信息系统深度融合。地理信息系统(geographic information system, GIS)是一种为了获取、存储、检索、分析和显示空间定位数据而建立的计算机化的数据库管理系统^[27]。GIS技术在海洋环境信息系统中具有非常广泛的应用。

2) 对数据存储、处理能力的要求较高。海洋环境信息产品的生产通常是典型的数据密集型+计算密集型应用,例如海洋环境预报、再分析产品制作等都需要动用大量的基础环境数据,其高时效性要求又需要强大的运算能力,对存储设备和计算设备提出了非常高的要求。大型数据中心、高性能计算机对任何涉海单

位而言都是一项巨大的成本支出。

3) 交叉应用日益增多。早期的海洋环境信息系统主要针对某项特定应用,其数据源、数据处理方法、服务对象等都较为单一。随着海洋领域的不断发展,具有交叉应用的海洋环境信息系统日益增多,例如地理、水文、气象等环境基础数据已经广泛应用于渔业管理、海洋经济预测、航道港口保障等方面。

经过多年的建设,中国建成了大量的专业海洋环境信息数据库,搭建了多套应用于海洋多种业务的信息系统。这些业务化运行的信息系统为中国海洋行政管理、海洋权益维护、海洋防灾减灾、海洋科学研究等方面提供了有力的信息保障和技术支撑。但这些各自

独立且内部结构相似的海洋信息系统也存在着如下问题:资源占用多、运行成本高;传统模式难以适应业务部署的快速要求、缺乏统一部署计算资源的规划;业务系统稳定性和可靠性低、系统维护难度大。因此,利用云计算技术,可以有效解决上述问题,提高海洋环境信息资源的可重用性与共享性。

2.2 海洋环境信息云计算的体系架构

综合考虑云计算的基本结构和海洋环境信息系统的特点,结合现有业务的实际情况,从3个层面设计海洋环境信息云计算体系架构,包括资源层、平台层和应用层(图3)。

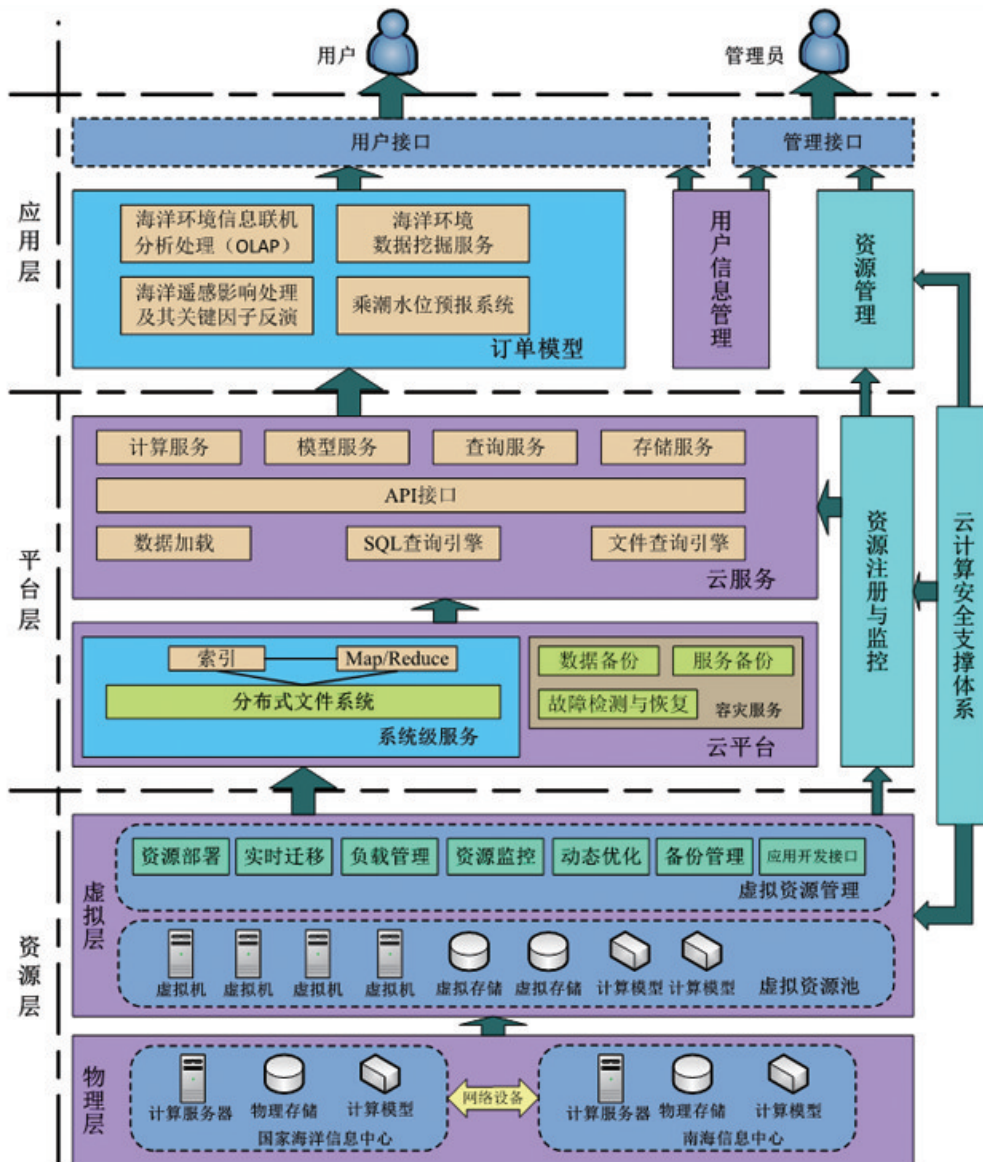


图3 海洋环境信息云计算体系架构

Fig. 3 Marine environment information cloud computing architecture

1) 资源层。

资源层作为整个系统的基础,主要功能是为平台层提供计算、存储和数据服务的资源保障。资源层进一步划分为物理层和虚拟层。

物理层由异地两个节点的服务器、存储设备组成,两个节点间通过互联网连接。

虚拟层通过虚拟化技术,将来自不同节点的、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;虚拟资源管理是虚拟化技术的重要方面,实现资源部署、资源监控、实时迁移、负载管理、动态优化、备份管理和应用开发接口等模块。

2) 平台层。

平台层是衔接资源层和应用层的中间件,包含云服务、云平台和资源注册与监控3个部分。

云服务是通过数据加载、结构化查询语言(structured query language, SQL)查询引擎和文件查询引擎对应用层提供计算服务、模型服务、查询服务和存储服务等基本服务,同时为用户提供应用程序编程接口(application programming interface, API),可以在满足云计算平台计算模型标准的基础上,开发自定义服务。

云平台包含系统级服务和容灾服务2个部分。系统级服务的功能是以分布式文件系统为基础,利用索引和MapReduce技术对用户提出的任务请求进行管理和调度。容灾服务包括数据备份、服务备份和故障检测与恢复。

资源注册与监控对资源层中的各种资源进行注册,为云服务提供资源索引服务;同时监控所有资源的使用情况,为应用层资源管理提供支持。

3) 应用层。

应用层作为整个系统的顶层,通过用户接口为用户提供交互界面,并在这个界面上提供订单式服务,同时通过管理接口为管理员提供管理界面,管理用户的权限等信息以及管理系统中的所有资源。

支撑用户接口和管理接口的是订单模型、用户信息管理和资源管理。其中订单模型整合多种海洋环境信息应用模块,并为用户提供使用上述应用模块定制个性化订单的功能。

用户信息管理保存了用户的相关信息和个性化设置,并为管理员提供对用户注册、权限的管理。

资源管理调用平台层中资源注册与监控模块,为管理员提供各种标准化、流程化的资源管理功能。

2.3 海洋环境信息云计算的标准体系

由于海洋环境信息云计算需要将大量异构的硬件、软件、数据等资源整合,并为用户提供标准化的服务,因此在云计算的实践中,建立相关的标准体系是十分重要的。

1) 软件服务标准。

软件服务标准是围绕用户,并提供海洋环境信息服务。

海洋信息管理服务标准包括数据安全、用户权限管理、版本控制管理、目录服务、元数据管理、访问粒度管理和收费管理等方面的内容。

数据生产服务是为用户提供数据生产、质量检查、数据更新等方面的服务;数据存储检索服务包括时态数据管理、海洋环境数据组织和检索、导航等服务;海洋信息的处理服务包括数据协同编辑、拓扑建立、数据转换、信息提取、制图、影像理解与合成以及要素综合服务;海洋信息增值服务是利用数据融合与数据挖掘技术提供决策支持服务,如海洋潮汐预报、航海路线导航等。

由于不同服务标准对服务质量的要求内容不同,所以对各种服务也有相应的服务质量标准的定义。系统提供统一的资源目录服务给终端用户,达到快速定位用户需要的资源。

2) 平台服务标准。

平台服务标准包括平台服务水平标准和平台服务评级标准。平台服务水平标准通过权限管理,在有效地保护海洋环境信息提供者权益及敏感信息安全的前提下,最大限度地为用户提供服务。平台服务评级标准针对不同的服务定义不同的服务级别,通过建立服务水平监督体系,保证服务达到规定的水平等级,如海洋环境信息数据集的完整度、比例尺等达到标准要求。

3) 私有云标准。

海洋环境信息云计算采用私有云结构,在组织内部管理中,不承受网络带宽和可用性问题和潜在安全风险的负担,为用户提供更好的控制、安全性和恢复能力,更好地处理数据保护和服务级问题。

私有云标准包括以下9个方面:确保严格的用户管理和业务定制规范;确保严格的身份认证;确保数据加密存储;确保数据的高可靠性;确保操作有记录;确保系统安全可靠;确保弹性扩展;确保接口和数据规范等有统一的标准;确保数据中心的有效管理。

4) 互操作性标准。

海洋环境的云计算需要在多个平台之间迁移应用程序并实现互操作,因此要制定一种具备互操作性的IT管理标准,用于描述消息的格式和消息交换的规则。互操作性标准包括3种:文档格式的互操作标准、应用编程接口(API)协议、通过权限机制控制不合法操作。

5) 模型与信息管制规范。

海洋环境信息云计算体系框架采用基于角色的访问控制模型,根据安全策略划分角色对每个角色分配操作许可,为用户指派角色,用户可通过角色间接地进行信息资源访问。

2.4 海洋环境信息云计算的安全体系

云计算中包含大量的资源共享及远程资源访问,因此必须建立完善的安全体系,具体包括以下7个方面^[28]。

1) 身份认证技术。云计算环境下,很多海洋环境信息应用要求用户可以跨越相对独立的安全域和安全地,访问各种海洋环境数据或者其他系统,因此,需要建立严格的身份管理、安全认证与访问权限控制,提供用户访问记录,做到访问可溯源^[29]。

2) 无边界网络的安全访问。云计算环境下,所有的用户连接到统一的云计算环境中,使用标准浏览器实现对各类海洋环境数据资源或模型资源的虚拟使用,用户不需要知道具体存放位置和访问口令,也不需要访问控制重新设计、调整,以及提供端到端的防护控制措施。

3) 统一的安全管理。通过融合传统的网络管理与安全管理的技术架构、功能模块与运维方式,使用户通过一个统一的Web管理控制台就能够对各类资源进行监控和安全事件管理,将网络管理与安全管理的日常运维工作进行整合,以简化用户管理的复杂性,降低网管信息重复采集带来的额外网络负担。

4) 虚拟化的安全。虚拟化是云计算的标志之一,但虚拟化导致应用进程间的相互影响更加复杂、数据的隔离与清除难以衡量、传统的分域防护难以实现,因此,安全技术必须支持虚拟化环境,能够伴随虚拟平台进行迁移。

5) 海洋信息数据多副本容错。云计算环境需要能够屏蔽单一存储设备的故障,以保证数据安全。目前,一般使用独立磁盘冗余阵列(redundant arrays of inde-

pendent disks, RAID5)技术实现数据容错。利用RAID5技术可在一个磁盘数据发生损坏时,利用剩下的数据和相应的奇偶校验信息去恢复被损坏的数据。

6) 站点故障处理与恢复。在服务器出现故障时,采用多机热备技术为用户提供持续可靠的服务。多机热备技术可以是实际的物理服务器也可以是虚拟的服务器,是通过即时备份数据的方式,实现系统的高可用性。由于海洋环境信息数据的海量性,采用共享存储器方式,无需考虑数据同步,不会产生数据不一致问题。

7) 应用程序容错。建立应用程序管理器,定期收集应用程序缓存数据、各种表及正在进行的进程等信息。如果应用程序出现故障启动保存有检查点信息的冗余应用程序。

3 资源虚拟化与海洋环境信息云平台构建技术研究

3.1 海洋环境信息云平台关键技术

基于上述研究与现有资源建设海洋环境信息云平台,需要重点解决两大问题:一是充分整合现有的各类资源,主要包括现存的存储设备、计算设备和大量的海洋环境信息数据,即解决资源虚拟化的问题,包括存储虚拟化、服务器虚拟化和海洋环境信息虚拟化;二是充分利用整合后的海洋环境信息数据,即解决云计算环境下海量数据的快速查询问题,本研究采用MapReduce技术,实现对海洋环境信息数据的高效查询。

3.2 存储虚拟化

存储虚拟化是指为物理的存储设备提供一个抽象的逻辑视图,可通过统一逻辑接口访问被整合的存储资源。目前较为主流的是基于存储区域网络(storage area network, SAN)的存储虚拟化技术^[30];根据实现方式的不同,可以分为主机级、存储设备级和存储网络级3个层次^[31]。

基于主机的虚拟化是将虚拟化层放在SAN中的应用服务器上,通过改造操作系统或者通过在服务器操作系统中嵌入或添加虚拟层来实现设备虚拟化。该方法不需增加硬件,只需安装具有虚拟化功能的软件模块,以驱动程序的形式嵌入到应用服务器的操作系统中,以逻辑卷的形式将分布在多机上的物理存储设备映射为统一的逻辑虚拟存储空间。

基于存储设备的虚拟化是将虚拟化层放在存储设备的控制器上。该方法直接面对具体的物理设备,性能较好。由于该虚拟化逻辑被集成到设备内部,存储虚拟化的管理简单方便、对用户透明。

基于存储网络的虚拟化是通过在SAN互联这一级增加一个元数据服务器,或者采用智能化路由器、交换机等来实现虚拟化的工作。其中,以专用服务器作为虚拟化平台,在其中存放整个存储系统的元数据,相关访问都以虚拟化平台为核心的方式,更加适用于基于策略和QoS的存储管理。

目前,现存的存储设备采购于不同时期,其品牌、性能和结构不尽相同,所有的存储设备中均存储有大量的海洋环境信息数据,而且正在被不同的业务系统所使用。如果采用基于主机的虚拟化,逐一改造存储服务器的操作系统,虽然不需要增加设备,但需要解决设备异构的问题,且改造工作量较大,有影响业务系统运行的技术风险。如果采用基于存储设备的虚拟化,虽然对单一设备的管理较为简便,但需要解决不同设备间的一致性,对云平台的管理不利。采用基于存储网络的虚拟化虽然需要增加专用服务器,但对现存存储设备及现有业务系统的影响最小,后期维护也较为方便。

综上所述,根据海洋环境信息云计算中存储设备的异构特性以及分布式海量存储需求,综合考虑对现有存储设备的整合利用的合理性及实现维护的方便性,拟采用基于专用服务器的SAN虚拟化技术,结合IDE/SATA磁盘虚拟化技术,从而实现存储的虚拟化(图4)。

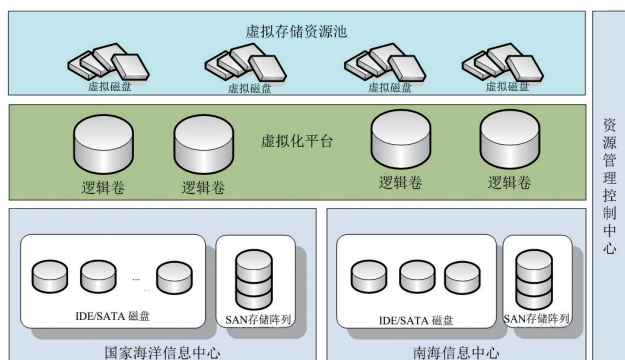


图4 海洋环境信息存储虚拟化框架

Fig. 4 Framework of virtualization of storage

3.3 服务器虚拟化

服务器虚拟化利用虚拟化软件向上提供对硬件的

抽象和虚拟化服务器管理,涉及到CPU虚拟化、内存虚拟化、设备与I/O虚拟化及虚拟机实时迁移等核心技术。目前有两种实现服务器虚拟化的方法^[32]:一种为寄宿虚拟化技术,通过在宿主机操作系统之上运行虚拟监视器来实现,其性能较低;另一种通过在硬件上运行虚拟化平台来实现服务器的虚拟化,其具有较高的性能。

海洋环境信息系统通常需要对大量的数据进行处理,对计算资源的需求较高,而目前计算资源已经是海洋环境信息系统的瓶颈资源。因此,海洋环境信息云平台拟采用虚拟化平台技术方案(图5)。在所有纳入云平台的计算服务器上均运行虚拟化平台系统,可以将全部计算资源整合为统一的计算资源池。

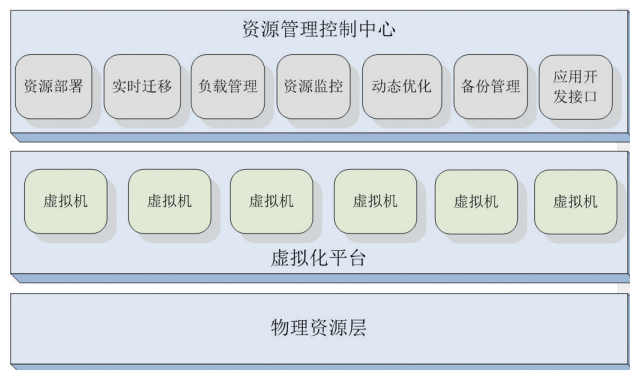


图5 服务器虚拟化框架

Fig. 5 Framework of virtualization of the servers

虚拟化平台直接实现对硬件资源的抽象处理,为上层服务提供了统一的指令集和设备接口,将一个物理服务器上的处理器、内存、存储器和网络资源抽象为多个虚拟机,实现服务器资源的虚拟化。同时,可以支持虚拟机的动态创建、迁移和退出,最大程度上满足了应用的灵活性。

资源管理控制中心主要提供对大量物理资源和虚拟资源的集成管理,由资源部署、动态迁移、负载管理、资源监控、动态优化以及应用开发接口等功能组成。

3.4 海洋环境信息虚拟化

海洋环境信息资源虚拟化主要是通过提供统一的虚拟海洋环境信息资源目录服务来实现,从而屏蔽底层各种海洋环境数据资源和海洋模型资源的多源、异构特性,以及在数据格式和存储方式上的巨大差异,为上层海洋模型服务和海洋环境数据服务提供统一的标

准化接口,简化访问海洋信息数据资源和海洋模型资源的复杂性(图6)。

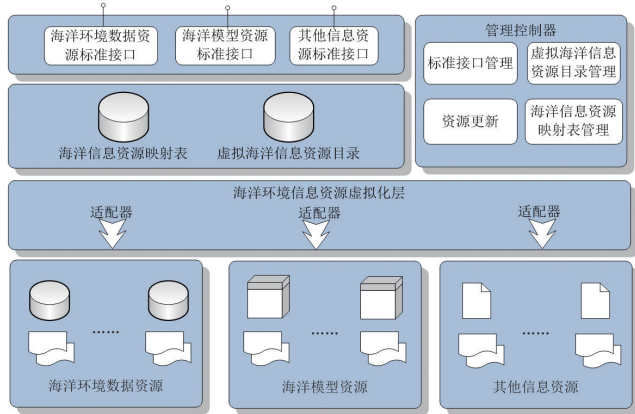


图6 海洋环境信息资源虚拟化框架

Fig. 6 Framework virtualization of marine environmental information resources

接口层为上层云计算平台和云服务提供标准统一的调用接口和访问接口,主要有海洋环境数据资源标准接口、海洋模型资源标准接口、其他信息资源标准接口等。海洋环境信息映射表主要用于存放虚拟海洋环

境信息资源和物理海洋环境信息资源之间的映射关系,而虚拟海洋环境信息资源目录主要存放各类虚拟海洋环境信息资源。管理控制器实现对标准接口、海洋信息资源映射表和虚拟海洋信息资源目录的构建、管理和维护操作,对虚拟资源目录信息的统计和海洋信息资源的更新等功能。海洋信息资源虚拟化层主要实现各类海洋信息数据资源、海洋模型资源以及其他海洋信息资源的虚拟化,并提供虚拟海洋信息资源与物理海洋信息资源之间的映射和转换功能,实现对物理海洋环境信息资源的透明访问。各类适配器实现对物理海洋信息资源的存取任务,并屏蔽底层差异。

3.5 基于MapReduce的海洋环境云平台技术架构

面对海量的数据信息,可以借助MapReduce分布式并行框架来提高大规模分布式系统上的查询处理效率,但是MapReduce对分片数据采取顺序子块读取的策略,需要把被处理的数据顺序地全部读取到内存中进行操作,没有使用相关的查询优化技术^[33]。因此,在海洋环境信息云计算中采用查询优化技术,以进一步提高数据密集型应用的数据处理性能(图7)。

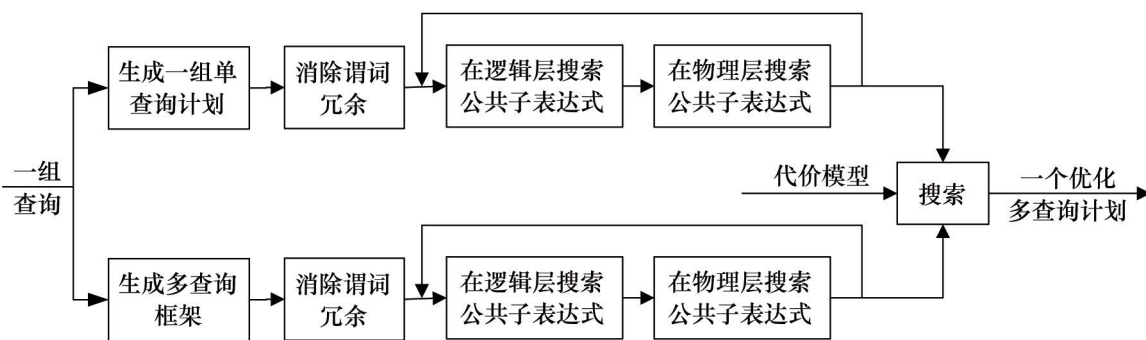


图7 多任务查询优化框架

Fig. 7 Framework of multi task query optimization

首先通过对查询相关数据的反聚簇等技术,将相关数据尽量分布到分布式系统的各个节点,以提高并行访问的效率;其次通过索引技术,提高查询数据的搜索效率和随机定位访问性能。

此外,海洋环境信息应用中,系统要支持大量的用户并行访问查询,会存在众多用户查询等待执行,不同的查询任务访问相同数据表或基于相同表做相似操作的概率较高。因此,在海洋环境信息云计算中,采用判定任务的谓词间关系、判定和提取公共子表达式、重写查询、查询结果数据缓冲机制等策略来实现多查询任

务优化。

1) 传统数据库查询算子的MapReduce转化实现。

由于MapReduce本身并不支持高效的结构化数据的查询和分析,因此需要实现选择、投影、连接、聚集和各种集合操作的MapReduce实现机制。一方面需要根据不同数据连接的特点,对MapReduce进行嵌套的操作,并实现对不同MapReduce操作流程结果的连接;另一方面需要根据具体应用的需要,来实现对数据仓库分析、数据挖掘和流关系数据原子操作的MapReduce机制。

2) MapReduce 集群的查询优化。

针对 MapReduce 并行架构未能使用查询优化技术的问题,在海洋环境信息云计算中将采用提高查询的并行度和使用索引技术相结合来进行查询优化。首先,将相关数据尽量分布到分布式系统的各个节点,以提高并行访问的效率;其次,采用海量数据的索引技术,优化查询执行计划,从而提高查询数据的搜索效率和随机定位访问性能。

4 海洋环境信息云平台示范系统的实践

4.1 物理结构设计

根据海洋环境信息云平台示范系统的设计目标和资源分布情况,示范系统由国家海洋信息中心和南海信息中心构成,提供数据资源、计算资源和存储资源(图8)。

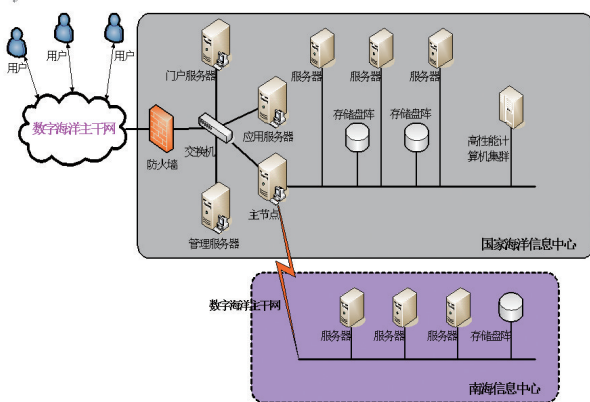


图8 原型示范系统的物理结构

Fig. 8 Physical structure of demonstration system

国家海洋信息中心作为云平台主节点,提供了20台服务器和1台24节点DELL集群作为计算资源,2台总容量为150 T的EMC磁盘阵列作为存储资源。

示范系统的核心计算资源以7台IBM X3650 M3服务器及2台IBM X3650 M2服务器组成,用于模拟云环境下的设备多样性。IBM X3650 M3服务器的配置为:2个Intel Xeon E5620处理器、2.4 GHz主频、12 M高速缓存、12 GB DDR3 RDIMM内存、热插拔SAS硬盘300G*3、SR M1015阵列卡、支持RAID 0/1/10/5及集成双千兆网卡。IBM X3650 M2服务器的配置为: Intel Xeon X5570 2.93 GHz处理器、标配6 GB DDR3-800内存。另外作为外围辅助计算资源的DELL PowerEdge

1850高性能集群,包括了48个CPU,理论峰值浮点计算能力为268.8Gflops, I/O节点和管理节点分别为2台DELL PowerEdge 2850。高速数据交换是由1台Myrinet光纤交换机和2台千兆网络交换机承担。DELL PowerEdge 1850服务器采用:2*1.4 GHz CPU处理器、2 GB DDR RAM内存、73 GB热插拔SCSI硬盘。此外,10余台IBM X3650 M2服务器作为辅助计算资源。

示范系统的核心存储资源主要采用EMC VNX5300磁盘阵列,这是一种高性能的iSCSI存储区域网络(SAN),最多可整合16台完全冗余的服务器,并可扩展至支持最高达90 TB的数据容量。这款机型具有双控制器,每个控制器含2个1 Gb以太网口,共有4个1 Gb以太网口。云平台外围辅助存储资源包括1台EMC Symmetrix DMX-4(120 T)和1台EMC CX3-120(15 T)。示范系统的计算能力达到0.5 T flops,存储能力超过150 T。

海洋环境信息云计算体系框架具有跨平台性(兼容Windows和Linux平台)、可扩展性(在国家海洋信息中心和南海信息中心两个计算节点的基础上,支持计算节点的增加;支持计算资源和存储资源的动态扩展)、高可用性(系统7×24 h运行,同时支持最多50个用户同时访问,故障率小于1%,平均故障恢复时间小于1 h);采用国际认同的技术标准和架构。系统提供150 T的存储能力和200 GHz的计算能力,支持对服务器、集群、SAN及海洋环境信息等软硬件资源的动态透明访问。系统对用户屏蔽底层软硬件资源的部署方式和数据资源的组织方式,用户使用的资源可根据实际需要由系统实现跨物理节点的可伸缩调配。

4.2 软件结构设计

海洋环境云平台示范系统实现以下4类软件:虚拟化软件、云平台支撑软件、资源和用户管理软件和应用服务软件(图9)。

虚拟化软件负责将各种物理资源虚拟为逻辑资源,并实现相应的管理控制功能。虚拟化软件的客户端部署在国家海洋信息中心节点和南海信息中心节点的所有数据库、作为存储资源的盘阵和作为计算资源的服务器上,管理端部署在国家海洋信息中心节点的管理服务器上。

云平台支撑软件基于Hadoop进行开发,负责利用虚拟的逻辑资源,形成云平台的基础云计算与云服务能力。云平台支撑软件的客户端部署在国家海洋信息

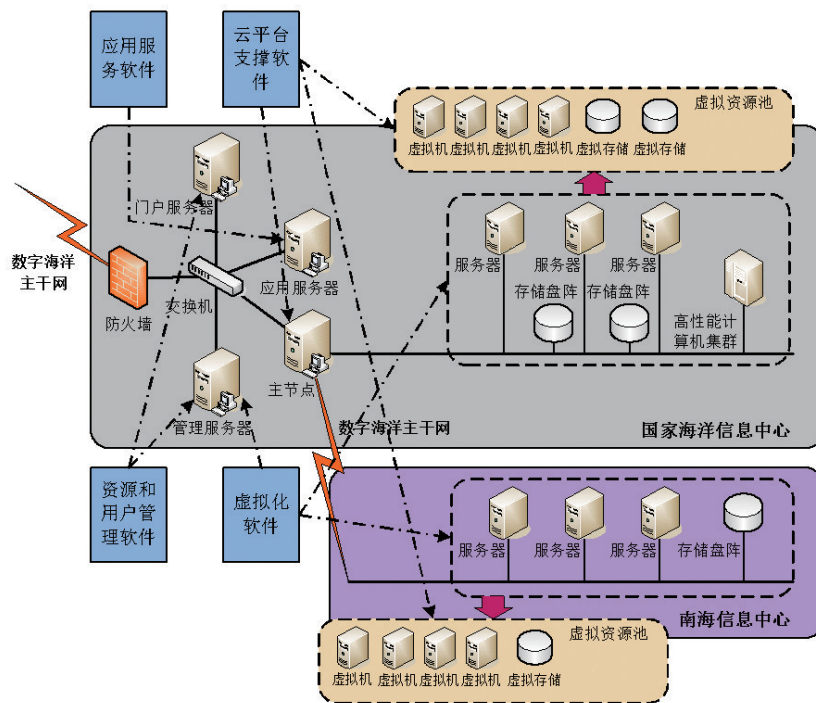


图9 示范系统的软件部署

Fig. 9 Software deployment of demonstration system

中心节点和南海信息中心节点所有由虚拟化软件虚拟出的虚拟服务器上,管理端部署在国家海洋信息中心节点的主节点服务器上。

资源和用户管理软件负责对各种物理资源进行管理,同时负责示范系统的用户管理。其中资源管理软件部署在国家海洋信息中心节点的管理服务器上,用户管理软件部署在国家海洋信息中心的门户服务器上。

应用服务软件既包含预先开发的用以实现各种海洋环境信息云计算典型应用的应用服务软件,也包含示范系统投入运行后,由用户定制开发的各种应用服务软件。应用服务软件均部署在国家海洋信息中心节点的应用服务器上。

示范系统的所有系统软件均基于C++或Java语言开发,所有为该系统开发的应用服务软件均需提供与C++或Java语言编程环境兼容的接口。

4.3 海洋环境信息云平台示范系统典型应用服务

1) 海洋环境信息联机分析处理(on-line analytical processing, OLAP)。

实现了基于Hive的OLAP分析系统,并借助Hadoop的MapReduce并行化技术实现了相关OLAP查询并行化执行。传统单机海洋OLAP与传统单机海洋

OLAP相比,当数据量快速增加时,能够有效降低系统查询的响应时间,而且随数据量增加依然保持了线性的并行性能,极大提高了系统多维查询的效率(图10)。

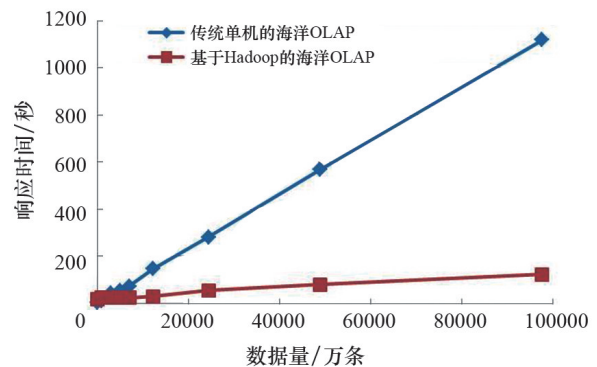


图10 海洋环境信息联机分析处理效果

Fig. 10 Performance of marine environment OLAP

2) 海洋环境数据挖掘服务。

实现了基于关联规则发现和异常模式发现2种数据挖掘方法。这2种数据挖掘方法的算法虽然精度较高,但难以处理大数据集。利用MapReduce机制,设计并行挖掘算法,能够有效地提高这2种数据挖掘方法的适用范围。在本系统中,以海洋观测网(array for real-time geostrophic oceanography, Argo)数据作为研究对

象,可发现温度和盐度的关联规则,以及其中的异常模式。用户上传数据文件后,Hadoop将数据分散存储于不同的节点中,利用MapReduce进行并行化处理后,结果也将以文件形式保存在云平台中,服务的用户交互界面如图11所示。



图11 海洋环境数据挖掘服务
Fig. 11 Marine environment data mining service

3) 海洋遥感影像处理及其关键因子反演。

针对海洋遥感影像处理的海量数据、复杂计算算法,各个计算节点之间采用数据划分的方式,每个节点处理不同的数据块;计算节点的内部,采用任务并行的方式,加快复杂算法的处理效率。实现从并行读写操作、数据分块、任务分解的多层次并行方式,以及云计算平台的遥感数据高效能处理。具体应用包括基于云平台的遥感影像融合、海岸线提取、SST反演、叶绿素反演等。叶绿素反演应用,采用的数据是存储于云平台上分辨率为1 m的高分辨率影像,当用户提交任务时,根据选定区域任务量提取相应的栅格数据,利用MapReduce技术,由任务调度中心将数据和任务分配到各个节点并行计算,最后基于空间索引机制实现结果的归并(图12)。

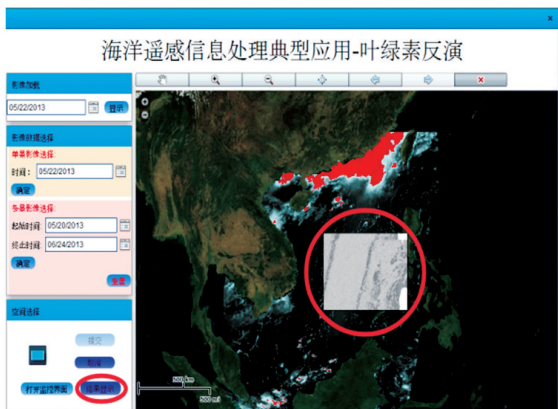


图12 叶绿素反演应用示意
Fig. 12 Application of chlorophyll inversion

4) 云计算环境下乘潮水位预报系统。

在云平台上基于Spark框架实现了乘潮水位分布式存储与并行计算,以及基于Spark框架的乘潮水位任务调度和Web交互系统。借助Spark框架解决传统计算模式成本较高、交互性、伸缩性差的问题。用户通过Web交互系统选取需要进行预报的港口后,Spark分布式计算框架提取出边界数据,利用Hadoop机制控制任务分配到各个节点,再控制各个节点运行计算过程,最后收集、汇总计算结果,上传到HDFS,如图13(a)。此外,用户可以通过Web交互系统,查询指定港口的潮位预报结果,如图13(b)所示。



图13(a) 乘潮水位预报用户界面
Fig. 13(a) User interface for forecast of tide riding level

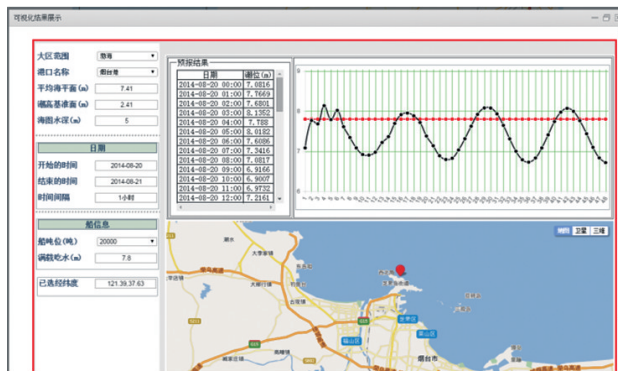


图13(b) 乘潮水位预报结果示意
Fig. 13(b) Result for forecast of tide riding level

5 结论

建设海洋环境信息云平台能够实现对现有硬件及数据资源的整合、降低信息系统成本、推动海洋环境信

息的共享和海洋产品的开发应用。本文在对国内外云计算发展状况的基础上,总结了云计算的基本结构和关键技术,对海洋环境信息云计算的体系架构进行分析,并针对资源虚拟化、云平台构建等关键技术进行研究;在此基础之上,开发建设了海洋环境信息云平台示范系统。经过在云平台上应用服务的实践,验证了云计算在海洋环境信息领域的应用效果,为继续整合各类资源、建设更为完善的海洋云平台、加快云计算技术在海洋信息领域的应用、建设覆盖范围更广的海洋综合信息服务系统等方面提供一定的借鉴。

参考文献(References)

- [1] Pallis G. Cloud computing: The new frontier of internet computing[J]. IEEE Internet Computing, 2010, 14(5): 70-73.
- [2] Armbrust M. A view of cloud computing[J]. Communications of the ACM, 2013, 53(4): 50-58.
- [3] Dikaiakos M D. Cloud computing: Distributed internet computing for it and scientific research[J]. IEEE Internet Computing, 2009, 13(5): 10-13.
- [4] Caithness N, Drescher M, Wallom D. Can functional characteristics usefully define the cloud computing landscape and is the current reference model correct?[J]. Journal of Cloud Computing, 2017, 6(1): 10.
- [5] 赵艳玲, 李战宝. 云计算及其安全在美国的发展研究[J]. 信息网络安全, 2011(10): 89-93.
Zhao Yanling, Li Zhanbao. Cloud computing and cloud security in the United States[J]. Netinfo Security, 2011(10): 89-93.
- [6] Kavitha V. A survey on security issues in service delivery models of cloud computing[J]. Journal of Network and Computer Application, 2011, 34(1): 1-11.
- [7] Ishida A. Cloud computing: Amazon EC2[J]. Ipsj Magazine, 2009, 50: 1068-1073.
- [8] Buyya R, Broberg J, Goscinski A. Cloud computing: Principles and paradigms[M]. Canada: Wiley Publishing, 2011.
- [9] 盖玉莲. 基于云计算的数据管理技术研究[J]. 电子设计工程, 2015(6): 71-74.
Gai Yulian. Research on cloud-computing-based data management technology[J]. Electronic Design Engineering, 2015(6): 71-74.
- [10] Naghshineh M, Ratnaparkhi R, Dillenberger D, et al. IBM research division cloud computing initiative[J]. IBM Journal of Research & Development, 2009, 53(4): 1-10.
- [11] Ramakrishnan R. Cloud computing at Yahoo[J]. International Conference on Management of Data, 2008, 6(4): 5-6.
- [12] Rhea Kelly. Open cloud consortium forms alliance with NOAA to release environmental data[EB/OL]. (2005-04-30)[2018-05-01]. <https://campustechnology.com/articles/2015/04/30/open-cloud-consortium-forms-alliance-with-noaa-to-release-environmental-data.aspx>.
- [13] 何宝宏, 李洁. 我国云计算发展的现状与展望[J]. 电信技术, 2012, 1(1): 22-24.
He Baohong, Li Jie. The present situation and Prospect of the development of cloud computing in China[J]. Telecommunications Technology, 2012, 1(1): 22-24.
- [14] 詹金艳. 互联网基础服务商世纪互联正式推出云计算产品—CloudEx[EB/OL]. (2009-04-16)[2018-05-01]. <https://www.csdn.net/article/1970-01-01/268783>.
Zhan Jinyan. Internet based service provider Century Internet officially launched cloud computing products—CloudEx[EB/OL]. (2009-04-16)[2018-05-01]. <https://www.csdn.net/article/1970-01-01/268783>.
- [15] 解鹏飞, 隋伟娜, 陶冠峰, 等. 云计算与海洋环境监测[J]. 海洋环境科学, 2013, 32(4): 576-580.
Xie Pengfei, Sui Weina, Tao Guanfeng, et al. Cloud computing and marine environmental monitoring[J]. Marine Environmental Science, 2013, 32(4): 576-580.
- [16] Kavis M. Architecting the cloud: Design decisions for cloud computing service models (SaaS, PaaS, and IaaS) [M]. New York: John Wiley & Sons, 2014.
- [17] Vasileiadou E M, Ullrich S, Tamm G. Cloud computing definitions and approaches—levels of abstraction: IAAS, PAAS, SAAS[EB/OL]. [2018-05-01]. http://www.asperado.de/sites/default/files/CloudComputingDefinitionsApproaches_2011.
- [18] 陈康, 郑纬民. 云计算系统实例与研究现状[J]. 软件学报, 2009, 20(5): 257-268.
Chen Kang, Zheng Weimin. Example and research status of cloud computing system[J]. Journal of Software, 2009, 20(5): 257-268.
- [19] 刘鹏. 云计算(第2版)[M]. 北京: 电子工业出版社, 2011.
Liu Peng. Cloud computing (Second Edition)[M]. Beijing: Publishing House of Electronics Industry, 2011.
- [20] 韩德志, 李楠楠, 毕坤. 云环境下的虚拟化技术探析[J]. 华中科技大学学报(自然科学版), 2012(增刊1): 262-265.
Han Dezhi, Li Nannan, Bi Kun. Study of virtualization technology in cloud environment[J]. Journal of Huazhong University of Science and Technology(Natural Science Edition), 2012 (suppl 1): 262-265.
- [21] Neiger, Gil, Santoni A, Leung F, et al. Intel virtualization technology: Hardware support for efficient processor virtualization[J]. Intel Technology Journal, 2008, 10(3): 167-178.
- [22] 孙振正, 龚靖, 段勇, 等. 面向下一代数据中心的软件定义存储技术研究[J]. 电信科学, 2014, 30(1): 39-43.
Sun Zhenzheng, Gong Jing, Duan Yong, et al. Research on

- software defined storage technology for the next-generation data center[J]. *Telecommunications Science*, 2014, 30(1): 39-43.
- [23] 阚文泉, 黄秋兰, 陈刚. 桌面网格环境下虚拟化技术的应用研究[J]. *计算机工程*, 2017, 43(3): 11-17.
Kan Wenxiao, Huang Qiulan, Chen Gang. Application research of virtualization technology in desktop grid environment [J]. *Computer Engineering*, 2017, 43(3): 11-17.
- [24] Vohra D. *Apache HBase and HDFS*[M]. Berkeley: Apress, 2016.
- [25] R Lämmel. Google's MapReduce programming model[J]. *Science of Computer Programming*, 2008, 70(1): 1-30.
- [26] 王英梅, 刘闯. 国内外海洋资源环境信息系统研究现状与发展趋势[J]. *资源科学*, 1999, 21(6): 75-79.
Wang Yingmei, Liu Chuang. Achievements and challenges of marine information systems[J]. *Resources Science*, 1999, 21(6): 75-79.
- [27] 吴信才, 白玉琪, 郭玲玲. 地理信息系统(GIS)发展现状及展望[J]. *计算机工程与应用*, 2000, 36(4): 8-9.
Wu Xincan, Bai Yuqi, Guo Lingling. Development and prospect of geographic information system[J]. *Computer Engineering and Applications*, 2000, 36(4): 8-9.
- [28] 魏红宇, 曲海鹏, 刘培顺, 等. 海洋环境云计算安全防护支撑体系的研究与构建[J]. *中国海洋大学学报(自然科学版)*, 2011 (增刊1): 429-432.
Wei Hongyu, Qu Haipeng, Liu Peishun, et al. Research and construction of ocean cloud computing protection and support system[J]. *Periodical of Ocean University of China (Natural Science Edition)*, 2011(suppl 1): 429-432.
- [29] 冯登国, 张敏, 张妍. 云计算安全研究[J]. *软件学报*, 2011, 22(1): 75-87.
Feng Dengguo, Zhang Min, Zhang Yan. Study on cloud computing security[J]. *Journal of Software*, 2011, 22(1): 75-87.
- [30] 李树全, 吴跃, 陈志飞. SAN存储虚拟化研究[J]. *微电子学与计算机*, 2009, 26(1): 39-41.
Li Shuquan, Wu Yue, Chen Zhifei. Research on virtualization of SAN[J]. *Microelectronics & Computer*, 2009, 26(1): 39-41.
- [31] 张式勤, 金敏玉, 宋祺. 存储虚拟化的应用场景研究及测试[J]. *移动通信*, 2011, 35(2): 26-28.
Zhang Shiqin, Jin Minyu, Song Qi. Research and test of application scenario for storage virtualization[J]. *Mobile Communications*, 2011, 35(2): 26-28.
- [32] Václavík K. Comparison of two server virtualization platforms [J]. *Journal of Physics Condensed Matter*, 2012, 5(5): 123-128
- [33] Dean J, Ghemawat S. MapReduce: A flexible data processing tool[J]. *Communications of the Acm*, 2010, 53(1): 72-77.

Application of cloud computing for marine environmental information

HUA Yanning¹, ZHOU Xue¹, CHEN Yanyun¹, LU Feng², CHEN Gang¹, ZHANG Lei¹

1. National Marine Data and Information Service, Tianjin 300171, China

2. CETC Ocean Information Co., Ltd., Beijing 100041, China

Abstract Based on a review of the present situation of the domestic and foreign developments and applications of the cloud computing, and in view of the features of the marine environment information system, this paper studies the system structure and related technologies of the cloud computing for the marine environment information. A cloud platform of the marine environment is designed and implemented, and some practical service is provided as an example for typical applications, which shows how to follow the cloud computing technology and the application development in the field of the marine information in the future.

Keywords marine environmental information; cloud computing; cloud platform; virtualization ●



(责任编辑 卫夏雯)