

文章编号:1001-893X(2011)06-0021-04

# 一种海量军事信息的数据云管理系统<sup>\*</sup>

闫振东,高 斌

(中国电子科学研究院,北京 100041)

**摘 要:**采用云计算的思想与技术平台,针对军事数据海量和异构的特点,研究了网络环境下跨域、异构、动态、海量数据组织与服务提供等关键技术,构建了对海量军事数据云存储的管理系统。测试结果表明,该系统能较好地解决海量军事数据云的分布式存储、透明访问和并发处理等问题,为民用技术在军事上的应用进行了有益的探索。

**关键词:**军事信息;云计算;异构;海量数据;管理系统

**中图分类号:**TP31 **文献标识码:**A **doi:**10.3969/j.issn.1001-893x.2011.06.005

## A Data Cloud Management System for Mass Military Information

YAN Zhen-dong, GAO Bin

(China Academy of Electronics and Information Technology, Beijing 100041, China)

**Abstract:** Based on the thought and technique platform of cloud computing, according to the characteristics of military information, the key techniques in network environment are studied, including cross domain, heterogeneous, dynamic, mass data structure and service providing. A management system for mass military data cloud storage is established. Test result shows that the problems related with the mechanism of distributed storage, transparent access and intercurrent process of data are solved. It is also a beneficial exploration to apply civil technique to military field.

**Key words:** military information; cloud computing; heterogeneous; mass data; management system

## 1 引 言

云计算(Cloud Computing)是近两年提出的新概念,是分布式计算(Distributed Computing)、并行计算(Parallel Computing)和网格计算(Grid Computing)的进一步发展,或者说是这些计算机科学概念的商业实现。总的来看,云计算还处于一个起步阶段,主要有 Amazon、IBM 和 Google、Hp 等各自的计算平台<sup>[1-3]</sup>。

当前战场形势下,信息数量激增,甄别难度加大。首先,随着各类侦察卫星及空中、地面和海上侦察监视设备的广泛使用,信息化战场形成了一个以计算机网络为基础、以侦察监视卫星为主的陆、海、空、天一体化战场感知系统,大量信息源源不断地通

过这一系统涌入指挥员的视野。例如科索沃战争中,北约利用自身广泛的卫星通信网络每天传输 15 万条以上的信息,包括加密和非加密声音、数据和传真信息。指挥机构要在信息的海洋中快速、准确筛选出有价值的信息,这的确是一个严峻的挑战。其次,过去的战争由于受到技术条件的束缚,人们对战场信息的认识仅仅局限于情报,战场信息管理也就简化为战场情报管理。但是随着信息技术在战争领域的广泛应用,战场信息的内涵和外延远远超出了情报的范围,要想获得战争的胜利,仅仅掌握情报是不够的,还必须了解、掌握整个战场空间和作战相关的数据、指令、消息、信号等敌我双方的信息。战场信息内涵的多样化使得信息管理对象的范围急剧扩

<sup>\*</sup> 收稿日期:2011-03-04;修回日期:2011-05-03

大,作战双方力求“兵不厌诈”导致信息的准确性、有效性不易鉴别,信息管理的复杂程度极度增加。

云计算的提出,为实现军事信息大规模数据的动态存储和处理提供了一种新的解决方案。但是,目前的云计算服务和平台主要都是由大型的商业公司提供的,开源的项目和平台还比较少,而这方面的相关研究也还比较缺乏。此外,目前针对大规模数据处理的平台都是领域无关的通用平台,因此,针对军事应用领域或者某种特定的数据格式,可以进一步研究相应的优化方法,提高数据管理和利用的效率。

面向海量军事信息的数据云管理系统要解决海量军事数据云的分布式存储、跨域透明访问与并行处理问题。基于此,本文采用了云计算的基本思想和技术来探索解决海量军事信息的管理难题。

## 2 系统总体框架设计

如图 1 所示,本系统主要由 4 个模块组成:数据云存储模块、数据管理模块、可编程 API 模块、数据广域网备份模块<sup>[4-5]</sup>。

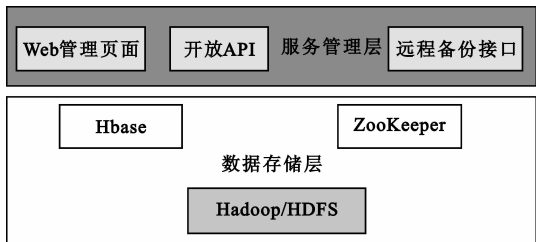


图 1 系统模块之间的组织关系

Fig. 1 The structure relationship of system module

### (1)数据云存储模块

该模块是数据中心的核心,实现数据云集群构建,数据存储的负载均衡、冗余复制,对卫星信息、雷达信息等各种军事数据进行结构化存储,并为管理模块提供节点动态注册的接口。

### (2)数据管理模块

该模块是数据中心的高层管理平台,实现对数据中心跨域的数据访问,以及对数据中心的透明访问、管理,实时监控数据中心状态。数据管理模块具体实现了对数据的可视化查询、透明存储数据、删除集群上的已有数据,同时可以监控数据中心的状态,如数据的剩余空间、集群节点状态、数据的冗余程度、集群中数据的分布、当前访问数据中心用户状态等信息。通过该模块,用户可以通过简单的操作,管

理、监视数据中心,对数据中心动态扩容<sup>[2]</sup>。

### (3)可编程 API 模块

该模块向开发用户提供了开放的 API 接口,这些接口的主要功能有数据查询、数据存储、冗余复制控制、map/reduce 应用模型等编程接口,用户可以通过调用这些开放的可编程接口进行二次开发。

### (4)数据广域网备份模块

该模块可以跨域地把一个集群节点的数据备份到其它节点上,当整个集群节点被破坏时,也可以由其它节点上备份的数据恢复该节点上的数据。

## 3 主要模块功能和结构分析

本系统主要由三大模块组成,分别是数据存储模块、数据管理模块和可编程 API 模块,下面分别详细介绍。

### 3.1 数据存储模块

对海量军事数据的存储,采用 google 的 GFS 文件系统模型,分布式存储数据,在开发过程中,通过对 apache 开源组织 Hadoop<sup>[6]</sup>项目的补充、修改,实现对军事数据的特殊支持。数据中心使用廉价的 Linux PC 机组成集群,在上面运行各种应用。即使是分布式开发的新手也可以迅速使用 Hadoop 的基础设施。分布式文件系统,隐藏下层负载均衡、冗余复制等细节,对上层程序提供一个统一的文件系统 API 接口。可以根据自己的需求对它进行特别优化,如超大文件的访问、读操作比例远超过写操作、PC 机极易发生故障造成节点失效等。

数据存储模块的结构如图 2 所示,主要由一个主节点和多个资源服务器构成。

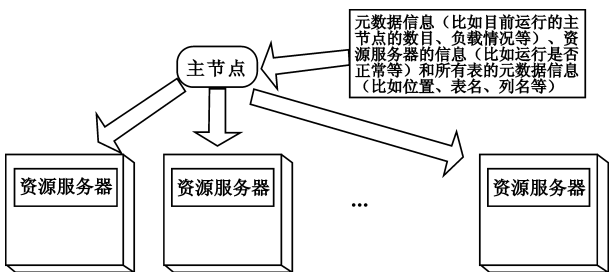


图 2 数据存储模块的结构图

Fig. 2 The structure of data storage module

### (1)一个主节点

主要用来管理主节点的元数据信息(比如目前运行的主节点的数目、负载情况等)、资源服务器的信息(比如运行是否正常等)和所有表的元数据信息

(比如位置、表名、列名等),同时还管理数据的分配、资源服务器的负载和增减以及垃圾数据的收集工作以及表元数据信息的变化,并和超级管理器交互,协同管理整个系统的正常运行。

### (2)多个资源服务器

资源服务器主要用来提供对该节点上数据的存取服务和管理,记录节点上被分配的任务情况,并向主节点提供节点的状态信息,如空余硬盘空间、内存使用情况等。

## 3.2 数据管理模块

数据管理模块实现了对数据的透明管理,用户可以通过 web 界面轻松地管理云数据系统,其功能主要包括可视化查询、存储数据、删除集群上的已有数据;同时可以监控数据中心的状态,如数据的剩余空间、集群节点状态、数据的冗余程度、集群中数据的分布、当前访问数据中心用户状态等信息。通过该模块,用户可以通过简单的操作管理、监视数据中心。

### (1)数据查询、存储、删除

根据用户的需要可以对云数据系统中的数据进行条件查询,并以表格的形式向用户展示查询的结果。通过 web 界面可以手动地向云数据管理中心加入任意数据,同时可以设置文件块大小、冗余复制方案等。当用户具有管理员权限的时候,就可以通过此管理界面删除云数据中心中的数据。

### (2)数据中心状态监控

此管理系统可以实时地监控整个集群的整体状态,如:集群节点组成、集群中各节点是否处于激活状态等。同时,通过管理系统还可以获得每个节点的具体物理资源信息,如 CPU 使用率、内存大小等,还可查看节点上存储块的信息等。以上状态都可以通过图形界面展现给用户。

### (3)冗余程度、数据均衡策略设置

冗余复制等级可以根据文件的要求,动态地设置冗余程度,系统默认的冗余是 3,也就是说每个块有 3 个备份块,分别存放在其它节点上。数据均衡策略设置相对比较开放,系统只提供一种简单的数据均衡模型,但用户可以自己编写数据均衡策略,然后通过管理软件的管理界面向系统提交方案,系统就会根据方案进行均衡管理。

### (4)数据节点动态注册

当用户希望加入新的节点时,可以把希望加入的主机名、主机 IP 等信息输入管理系统,系统会根据用户提供的信息增加数据节点,并注册到

主节点,实现数据节点的动态注册。新的数据节点加入系统后,系统会根据用户设定数据均衡策略进行数据均衡处理。

## 3.3 可编程 API 模块

面向海量军事信息的数据云管理系统提供可编程 API 的目的就是减小用户并行编程的困难,不必了解并行编程的内容,就可以对云存储中的数据进行并行化处理。

在系统底层采用 MapReduce<sup>[7]</sup>分布式并行计算模型。MapReduce 分布式计算模型能够很好地利用集群和多核的特点,方便地在普通计算机构成的集群中执行并行计算,同时也简化了并行计算模型,降低了并行程序设计的复杂度。本系统主要是利用了 Apache 的开源项目 Hadoop 提供的 MapReduce 支持,在其上执行并行计算、任务调度、负载平衡等工作。同时,系统提供了一个作业队列和一个优化的作业调度程序,包括数据的拷贝与传输、缓存中间结果、最小化平均响应时间等。

(1)采用开源的 Apache 项目 Hadoop 作为执行平台,提供 MapReduce 分布式并行计算支持,能够方便地部署在大规模的普通计算机构建的集群上,Hadoop 本身提供了负载均衡以及容错机制,降低了系统复杂度。

(2)系统保存一个等待执行的作业队列,作业之间的调度采用最小响应时间策略。

(3)区分对待信息。对于可靠的、经常使用的数据信息,将在大多数节点上保存该数据的拷贝,对于作业执行计划之间的数据冲突,采用了优化的数据复制和传输算法。

在 MapReduce 基础之上提供一些并行数据处理编程接口(API),其特点是通过使用简单的可用接口,能对复杂的数据处理流程进行操作,并对底层分布式处理细节,对并行化的计算和处理进行封装,简化用户的编程行为。图 3 说明了这一技术的具体实施。

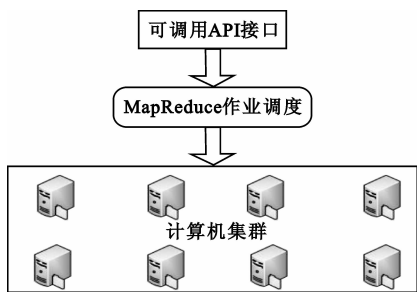


图 3 采用 MapReduce 的分布式计算模型  
Fig.3 The distributed module using MapReduce

## 4 结 语

云计算本来是一种民用商业计算模型,属于分布式计算技术。云计算模型多种多样,功能也各不相同,但基本思想是一致的。鉴于云计算在军事应用方面的欠缺,本文采用了这种理念和思想,又针对军事信息数据的特定需求和特点,提出了一种数据云管理系统。和通用的云计算平台不同,本系统功能性较强,不关注通用性而只针对军事海量信息的存储和处理。试验结果表明,云计算模型非常有助于解决大量军事数据的处理。本文对民用技术在军事上的应用进行了有益的探索和实践,两者的有机融合将会对我国信息技术的发展带来更大的益处。

但是,由于军事信息数据的实时性,以及战场环境的复杂性,导致对系统的性能需求也差别很大。鉴于定量指标的不确定性,对于云计算技术能否满足需求还有待后续的研究来证明。

## 参考文献:

- [1] 刘越. 云计算技术与应用[M]. 北京:工业和信息化部电信研究院通信信息研究所, 2009:3-10.  
LIU Yue. Cloud computing technology and application[M]. Beijing: Teleinfo Institute of China Academy of Telecommunication Research of MIIT, 2009:3-10. (in Chinese)
- [2] Zhang H, Goel A, Govindan R. Using the small world model to improve freenet performance[C]//Proceedings of the 21st IEEE. New York:IEEE, 2002:555-574.
- [3] Bak P. How Nature Works: The Science of Self-Organized Criticality[M]. New York: Copernicus Inc., 1996.

- [4] 赵晓哲, 郭锐. 军事系统研究的综合集成方法[J]. 系统工程理论与实践, 2004(10):127-130.  
ZHAO Xiao-zhe, GUO Rui. The integration method of Military system research[J]. System Engineering Theory and Practice, 2004(10): 127-130. (in Chinese)
- [5] 韩家炜. 数据挖掘: 概念与技术[M]. 北京:机械工业出版社, 2007.  
HAN Jia-wei. Data Mining: Concept and technology [M]. Beijing: China Machine Press, 2007. (in Chinese)
- [6] Chuck Lam. Hadoop In Action[M]. American: Manning Publications, 2010.
- [7] 王晋川, 何宏, 张福临. MapReduce 框架与调度容错机制研究[J]. 中国储运, 2010(12):90-91.  
WANG Jin-chuan, HE Hong, ZHANG Fu-lin. MapReduce frame and schedule fault-tolerant research[J]. China Storage & Transport, 2010(12):90-91. (in Chinese)

## 作者简介:

闫振东(1981—),男,山西晋中人,2007年获硕士学位,现为工程师,主要从事信息网络、海量数据存储方面的研究工作;

YAN Zhen-dong was born in Jinzhong, Shanxi Province, in 1981. He received the M.S. degree in 2007. He is now an engineer. His research concerns information network and mass data storage.

Email: eagle-lion@163.com

高 斌(1974—),男,江苏江阴人,2003年获博士学位,现为高级工程师,主要从事数据处理、系统集成设计研究。

GAO Bin was born in Jiangyin, Jiangsu Province, in 1974. He received the Ph.D. degree in 2003. He is now a senior engineer. His research interests include data processing, and techniques of integration for system.