第35卷 第24期 计 算 机 工 程 2009年12月
Vol.35 No.24 Computer Engineering December 2009
·软件技术与数据库· 文章编号：1000—3428(2009)24—0084—03 文献标识码：A 中图分类号：TP311

# 基于MPI的云计算模型

郭本俊，王 鹏，陈高云，黄 健
(成都信息工程学院并行计算实验室，成都 610225)

**摘 要：**根据消息传递接口(MPI)的特点，提出云计算在MPI领域的应用方法，包括MPI的云计算算法设计模型、云计算原理、核心计算模式、处理流程，并介绍云计算的分布式及并行化特性。理论分析结果表明，该算法是有效可行的，优于传统并行技术，能够为算法分布化及并行化提供新思路。

**关键词：**云计算；消息传递接口；机群系统；Hadoop架构

# Cloud Computing Model Based on MPI

GUO Ben-jun, WANG Peng, CHEN Gao-yun, HUANG Jian
(Parallel Computing Laboratory, Chengdu University of Information Technology, Chengdu 610225)

**【Abstract】** According to the features of Message Passing Interface(MPI), the cloud computing application methods based on MPI, including the MPI cloud computing algorithm design model, cloud computing principles, the core model, and the process are proposed. The distributed characteristic and parallel characteristic are introduced. Theoretical analysis results show this algorithm is feasible, effective and superior to the traditional parallel technology, and it can provide the new method to distributed and parallelize the ordinary algorithms.

**【Key words】** cloud computing; Message Passing Interface(MPI); machine cluster system; Hadoop frame

## 1 概述

随着信息化处理需求的增长，普通计算机的计算和存储在一定程度上制约着现代化办公，人们希望利用一台笔记本或一部手机，通过网络服务实现所有需要，甚至包括超级计算这样的任务。传统的并行技术[1]已为人类的发展做出重要贡献，但仍无法满足当前日益增长的办公和科研的需求，新型的并行计算技术——云计算[2]应运而生。

2006年底谷歌推出了“Google 101计划”，并正式提出“云”的概念和理论，该思想将云计算这种新兴的共享基础架构方法、需求及应用体现出来，而Amazon, Google和IBM则是第1批将云计算引入公众视线的公司。

本文阐述基于消息传递接口(Message Passing Interface, MPI)的云计算模型，并依据MPI原理建立与之对应的处理算法，将MPI和云计算结合在一起，实现并行化云计算的MPI。

## 2 云计算

### 2.1 基本原理

云计算[3]是分布式处理(distributed computing)、并行处理(parallel computing)和网格计算(grid computing)的发展与延伸，也是这些计算机科学概念的商业实现。

云计算是种新兴的共享基础架构的方法，其基本原理是透过网络将庞大的存储和计算处理程序分布到大量分布式计算机中，并提供相应的应用程序服务，使得企业能将资源切换到需要的应用上，根据需求访问计算机和存储系统。

云计算是基于互联网的超级计算模式，通过架构一个分布的、可全球访问的资源结构，使数据中心在类似互联网的环境下运行计算，即把存储于个人电脑、移动电话和其他设备上的大量信息和处理器资源集中在一起，协同工作。

### 2.2 云计算的兴起及发展

云计算是种全新的领先信息技术，结合IT技术和互联网实现超级计算和存储能力，而推动云计算兴起的动力是高速互联网连接的发展、更加廉价且功能强劲的芯片及硬盘、数据中心的发展。

目前，Sun、IBM、微软、Google、Amazon等信息业巨头都已经参与到云计算研究和开发中。IT巨头Sun公司在2006年推出了基于云计算理论的“黑盒子”计划，目前已经进入发售阶段；蓝色巨人IBM命名了“蓝云”计划；微软全世界有数以亿计的Windows用户，通过Windwos Live提供云计算服务实现一般的设备存储转移到任何时间都可以存储的模式；互联网企业的先行者Google的搜索引擎可以视为云计算的早期产品，其开放式的平台体现了云计算模式的精髓，其云计算服务所需要的绝大部分基础软件都是开源的；互联网上最大的在线零售商亚马逊提供弹性计算云，为独立开发人员及开发商提供云计算服务平台。

## 3 MPI简介

### 3.1 MPI的概念

当前最流行的高性能并行体系结构中比较常用的并行编程环境[4]分为2类：消息传递和共享存储。消息传递并行处理开销比较大，适合于大粒度的进程级并行计算，相对其他并行编程环境，它具有很好的可移植性，几乎被所有并行环境支持；还具有很好的可扩展性，具有完备的异步通信功能，能按照用户要求很好地分解问题、组织不同进程间数据交换适合于规模可扩展性并行算法。

**基金项目：**国家自然科学基金资助项目(60702075)；成都信息工程学院发展基金资助项目(KYTZ200819, CSRF200701)；成都信息工程学院校选科研基金资助项目(CRF200911)

**作者简介：**郭本俊(1975－)，男，讲师、硕士研究生，主研方向：高性能并行计算；王 鹏，博士后；陈高云，副教授；黄 健，讲师

**收稿日期：**2009-09-10 **E-mail：**guobenjun@cuit.edu.cn

MPI 能广泛应用于多类并行机群和网络环境，是建立在多种可靠的消息传递库的基础上的一种接口模式。

MPI 是消息传递并行程序设计的标准，用于构建高可靠、可伸缩及灵活的分布式应用程序，如工作流、网络管理、通信服务客户服务、天气预报和供应链管理等系统。

### 3.2 MPI 函数

MPI 是个库，共有上百个函数调用接口，Fortran、C 语言及 C++可直接调用这些函数。MPI 中只需要掌握 6 个常用函数就能掌握 MPI 的基本功能，完成几乎所有的通信功能。表 1 为 MPI 的 6 个基本函数。

**表 1　MPI 基本函数**

| 函数 | 功能 |
|---|---|
| MPI_Init | 初始化 |
| MPI_Finalize | 结束 |
| MPI_Comm_size | 获取进程个数 |
| MPI_Send | 发送 |
| MPI_Recv | 接收 |
| MPI_Comm_rank | 获取进程标识号 |

### 3.3 MPI 的消息传递过程

MPI 是基于消息传递的并行编程环境，通过定义核心库程序的语法、语义使各个并行执行部分之间通过传递消息来交换信息、协调步伐、控制执行。

消息传递过程的并行程序设计流程如图 1 所示。

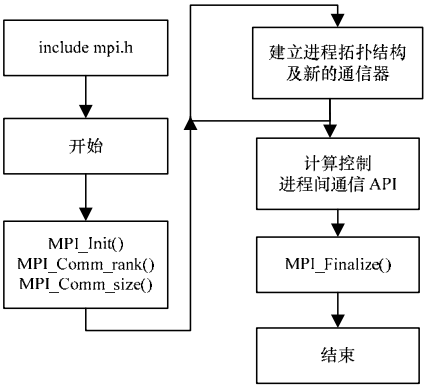


**图 1　MPI 并行程序设计流程**

在所有 MPI 程序中首先包含 mpi.h 头文件，然后由 MPI_Init()完成程序所有的初始化工作，并进入系统，进程调用需要用到的函数、应用程序，最后采用 MPI_Finalize()结束该进程。

### 3.4 并行机群 MPI

机群系统是由大量节点构成的，由点对点消息传递和组通信的函数实现，为编写并行程序提供方便，其中，组通信实现通信、同步和计算 3 个功能。

机群并行系统与传统的并行处理系统相比，机群大多采用商用工作站、PC 和通用 LAN 网络，其优点在于系统易于实现，节点主机及系统管理相对容易，且可靠性高。

## 4 MPI 云计算架构体系

### 4.1 MapReduce 模式[5]

Google 提出云计算的核心计算模式 MapReduce，是种分布式运算技术，也是简化的分布式编程模式，用于解决问题的程序开发模型，也是开发人员拆解问题的方法。

MapReduce 模式的思想是通过自动分割将要执行的问题(程序)、拆解成 Map(映射)和 Reduce(化简)的方式，其拆解过程的实质是将问题分为几个部分，划分技术[6]可以应用于程序的数据，将数据分解，然后对分解的数据并行操作，这样就将并行处理技术与云计算的核心技术融合在一起了。在此，采用一个简单而实际的数据划分实例说明，假设一个数列，$x0,x1,\cdots,xn$-1 共 $n$ 个数据项，将数列分为 $P$ 个部分，每个部分有 $n/p$ 个数据$(x_0\cdots x_{(n/p)-1}),(x_{n/p}\cdots x_{(2n-1)}),\cdots,(x_{(p-1)n/p}\cdots x_{n-1})$，采用 $P$ 个处理器分别对每一个分组进行处理。

在自动分割后通过 Map 程序将数据映射成不相关的区块，分配(调度)给大量计算机处理达到分散运算的效果，再通过 Reduce 程序将结果汇整，输出开发者需要的结果。MapReduce 的处理流程如图 2 所示。

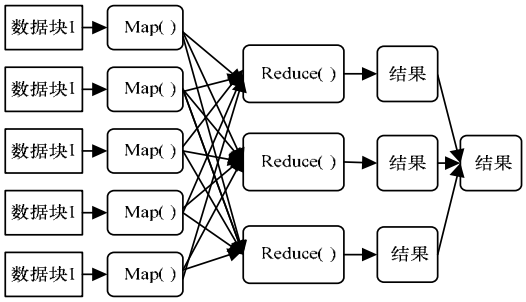


**图 2　MapReduce 的处理流程**

MapReduce 借鉴了函数式程序设计语言的设计思想，其软件实现是指定一个 Map 函数，把键值对(key/value)映射成新的键值对(key/value)，形成一系列中间结果形式的 key/value 对，然后把它们传给 Reduce(规约)函数，把具有相同中间形式 key 的 value 合并在一起。Map 和 Reduce 函数具有一定的关联性。其算法描述为

Map (k,v)-> list(k1,v1)

Reduce (k1,list(v1))->list(v1)

其中，v,v1 可以是简单数据，也能是组数据，对应不同的映射函数规则。在 Map 过程中将数据并行，即把数据用映射函数规则分开，而 Reduce 则把分开的数据用规约函数规则合在一起，即 Map 是个分的过程，Reduce 则对应着合。比如 Map((*)3 [2,5,9])表示其映射规则为乘 3，映射成为[6,15,27], Reduce((*)[2,5,9])规约的规则为求积，得到结果为 90。

MapReduce 模式的程序能在机群上实现并行化，常用于海量数据的并行处理，管理计算机之间必要的通信。

MapReduce 应用广泛，包括简单计算任务、海量输入数据、集群计算环境等，比如有分布 grep、分布排序、单词计数、Web 连接图反转、每台机器的词矢量、Web 访问日志分析、反向索引构建、文档聚类、机器学习、基于统计的机器翻译等。

### 4.2 Hadoop 架构

在 Google 发表 MapReduce 后，2004 年开源社群用 Java 搭建出一套 Hadoop 框架，用于实现 MapReduce 算法，能够把应用程序分割成许多很小的工作单元，每个单元可以在任何集群节点上执行或重复执行。此外，Hadoop 还提供一个分布式文件系统(HDFS)及分布式数据库(HBase)用来将数据存储或部署到各个计算节点上。Hadoop 框架具有高容错性，及对数据读写的高吞吐率，能自动处理失败节点。

在图 3 所示 Hadoop 架构中，MapReduce API 提供 map 和 reduce 处理，HDFS 分布式文件系统和 HBase 分布式数据库提供数据存取。

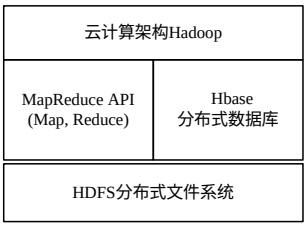


**图 3　Hadoop 架构**

### 4.3　MPI 云计算模型

用户将要执行的程序或处理的问题提交云计算的平台 Hadoop，其执行过程如图 4 所示。

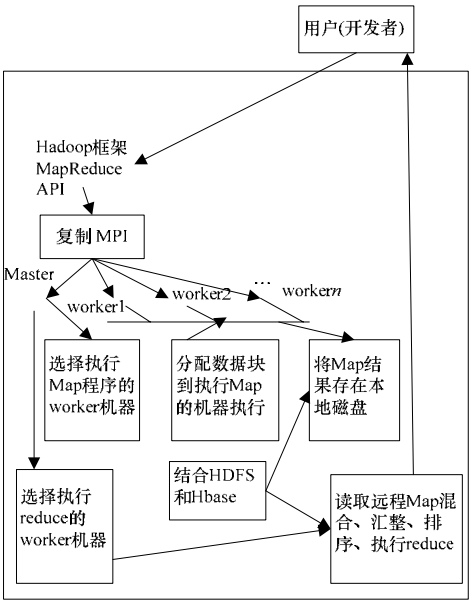


**图 4　MPI 云计算执行过程**

在该模型中，Map 尽量在数据输入近端执行，以避免数据大量移动。Map 映射把指令分发到多个 worker 上去，reduce 规约把 Map 的 worker 计算出来的结果合并，而 Master 则定期探测 worker，当出现故障时，重新执行 Map，reduce 若未完成则重新执行。在执行中任务粒度一般要求小数据块小于或等于 HDFS 中一个 Block 大小，且 Map 和 reduce 大于 worker 的数目以利于负载均衡和故障恢复。

### 4.4　算法并行化

在云计算实现中，借助 Hadoop 框架及云计算核心技术 MapReduce 实现计算和存储，其分布式文件系统 HDFS 和分布式数据库 HBase 及 map 算法充分体现云计算的分布式运算特性。

在消息传递处理中将 MPI 并行程序与并行机群系统及云计算框架和核心技术整合在一起，充分利用 MapReduce 的并行化和分布式计算来实现算法的分布式、并行计算和存储，达到高吞吐率、大规模数据运算能力。

## 5　应用和扩展

MPI 云计算算法在计算机应用领域得到广泛应用，如需要大量数据交换和计算的业务天气预报、天气模拟、泥石流模拟、星系模拟、网页索引数据库运算、垃圾邮件过滤、资料的全球移动、病毒分析、在线服务软件等方面。

目前云计算应用主要体现为 7 种类型：

(1)软件即服务(SAAS)。通过浏览器把程序提供给成千上万的用户应用，如在线软件服务。

(2)实用计算(utility computing)。将内存、I/O 设备、存储以及计算能力整合成一个虚拟的资源池，为整个网络提供服务。

(3)网络服务。网络服务提供 API 让开发者开发更多基于互联网的应用。

(4)平台即服务。提供开发平台服务，便于开发者用本地的设备来开发程序并通过互联网和其服务器传给用户。

(5)管理服务提供商(MSP)。面向 IT 行业，常用于邮件病毒扫描、程序监控等。

(6)商业服务平台。SAAS 和 MSP 的混合应用，为用户和提供商之间的互动提供一个平台。

(7)互联网整合。整合互联网服务类似的公司，方便用户对服务供应商的比较和选择。

随着对云计算技术开发和普及，会有越来越多的领域进入到云计算时代，实现计算和存储“云”的处理。云计算这种全新的商业模式必将引领全球新的产业革命，并对未来计算科学的发展提供新的发展思路。

## 6　结束语

本文提出云计算在消息传递、程序并行化上的应用，并给出云计算核心算法 MapReduce 及 Hadoop 架构模式，建立 MPI 的云计算理论基础，为机群系统分布式数据及海量数据的运算处理实现提供新思路。

云计算作为一种新兴的并行计算技术，在数据挖掘、图像处理、大规模数据处理等领域有着广泛的应用前景。对于云计算还需要人们去作更深入的算法开发和研究，将其应用于科研、办公和生活方面，以更好地发挥其作用。

### 参考文献

[1] 陈国良, 安　虹, 陈　崚. 并行算法实践[M]. 北京: 高等教育出版社, 2004.

[2] Thomas C. Google and IBM Partner to Push Cloud Computing[Z]. (2007-08-08). http://www.informationweek.com/news/internet/showArticle.jhtml?articleID=202400042.

[3] Stephen B. Google and the Wisdom of Clouds[Z]. (2007-12-13). http://www.businessweek.com/magazine/content/07_52/b4064048925836.htm.

[4] 何艳辉, 朱珍民. 基于消息传递并行计算环境[J]. 湘潭大学社会科学学报, 2003, 27(5): 233-235.

[5] Dean J. MapReduce: Simplified Data Processing on Large Clusters[C]//Proc. of the 6th IEEE Symposium on Operating System Design and Implementation. San Francisco, CA, USA: [s. n.], 2004.

[6] Barry W. Parallel Programming[M]. 陆鑫达, 译. 2 版. 北京: 机械工业出版社, 2005.

编辑　陈　文