

# Hadoop下水环境模拟集群运算模式

马金锋<sup>1</sup>, 唐力<sup>2</sup>, 饶凯锋<sup>1</sup>, 洪纲<sup>3</sup>, 马梅<sup>1,4</sup>

1. 中国科学院饮用水科学与技术重点实验室(中国科学院生态环境研究中心), 北京 100085;
2. 深圳市环境监测中心站, 广东 深圳 518057; 3. 石家庄市环境信息中心, 河北 石家庄 050051;
4. 中国科学院大学资源与环境学院, 北京 100190

## 摘要

水环境数值模型是模拟、分析及预测水体中物质迁移转化过程及其效应的有效工具。水环境模型的高性能批量计算是当前水环境模拟研究的热点。大数据技术中的分布式集群计算模式为水环境模拟批量计算提供一种可行的解决方案。探索了水环境数值模型在大数据分布式计算框架下的适应性,提出了一种适用于水环境模拟的大数据分布式集群运算模式,并通过实例验证了该运算模式的可行性。

## 关键词

水环境模拟 ; 集群运算 ; 大数据 ; Hadoop

中图分类号 : TP311

文献标识码 : A

doi: 10.11959/j.issn.2096-0271.2019051

## *Cluster computing mode for water environment simulation based on Hadoop*

MA Jinfeng<sup>1</sup>, TANG Li<sup>2</sup>, RAO Kaifeng<sup>1</sup>, HONG Gang<sup>3</sup>, MA Mei<sup>1,4</sup>

1. Key Laboratory of Drinking Water Science and Technology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China
2. Shenzhen Environment Monitoring Center, Shenzhen 518057, China
3. Shijiazhuang Environmental Information Center, Shijiazhuang 050051, China
4. College of Resources and Environment University of Chinese Academy of Sciences, Beijing 100190, China

## *Abstract*

Water environment numerical models are effective tools for the simulation, analysis and prediction of the processes of pollutant transport and transformation in water. The development of high-performance batch computation of water environment models has long been a hot topic. The distributed cluster computing mode based on big data technology is a promising approach for massive data management and batch computation, which provides a viable solution to large-scale water environment simulations. The adaptability of water environmental models under the framework of big data technology was explored, and a distributed cluster computing mode for water environment simulations was proposed. Moreover, the feasibility of adapting Delft3D model for cluster computing under Hadoop MapReduce environment was verified with real examples.

## *Key words*

water environment simulation, cluster calculation, big data, Hadoop

## 1 引言

水环境模型是指水体中的污染物(营养物质、悬浮物、藻类、有毒物质等)随水流在迁移过程中,因水动力和生物化学等因素的影响而发生的物理、化学和生物反应的数学描述和模拟<sup>[1]</sup>。水环境中污染物的迁移转化过程模拟和预测是水生态环境健康诊断、预测、预警及控制管理的重要基础,水环境数学模型在国内外已有非常广泛的应用,并且成为国内外学者研究的热点<sup>[2-4]</sup>。随着对水环境过程认知的进步、计算机运算能力的增强以及模型大量的普及应用,水环境模型取得了快速的发展<sup>[5-7]</sup>,功能日益强大,但模型也变得更加复杂,这对运算的需求急剧增加。

尽管过去几十年间计算机的运算速度和容量得到快速提升,但水环境模型的高性能批量计算一直面临巨大挑战<sup>[8]</sup>。综合运用并行技术和集群技术提高模型计算效率已成为环境领域和高端计算领域的研究热点<sup>[9]</sup>。目前,研究工作主要集中在CPU<sup>[10-12]</sup>、GPU<sup>[13]</sup>独立或协同<sup>[14]</sup>并行计算,包括通过单机多处理器实现高速计算能力、与普通计算机通过高速网络互联共享计算能力实现集群计算<sup>[8]</sup>。前者严重依赖于单机性能,成本昂贵且计算速度的提高空间有限;后者将任务分散到各个节点执行,适用于计算密集型的作业,当节点需要访问的数据量较庞大时,网络带宽可能会成为系统的性能瓶颈<sup>[15]</sup>。此外,在大规模分布式计算环境下,协调各个进程是一个很大的挑战,其中最困难的是合理处理系统的部分失效问题<sup>[16]</sup>。从本质上讲,上述并行和集群技术都是基于区域分解和分块计算策略<sup>[9]</sup>的,单算例被分解为多个子任务同时执行,以减少执行时间,适合单算例高性

能计算。

在水环境模拟应用领域,水环境模拟预测、水污染风险动态预警、水质目标管理、突发水污染事故应急决策、洪水风险控制图等应用对模型建模过程(参数率定、不确定性分析、模型验证)以及应用过程(情景分析)均提出了批量计算需求。单算例模式不适合批量计算应用。

大数据技术提供了一种新兴海量数据管理和计算模式,尤其是以Hadoop为代表的开源大数据平台,采用无共享(shared-nothing)框架,能够实现失败检测,并提供良好的横向扩展和容错处理机制。这种分布式集群计算架构隐藏了并发、容错、数据分布和均衡负载等细节,可以运行在一群廉价的PC上,为水环境模拟批量计算提供了一种潜在的、经济可行的解决方案。

本文针对多算例并行计算需求,探索水环境模型在大数据技术框架下的适应性,提出一种适用于水环境模型模拟的分布式集群运算模式,以解决目前集群计算面临的横向扩展和容错处理问题,即不但需要将计算“本地化”以节约网络带宽,从而获得高效率的计算性能,而且需要设计良好的机器故障处理(容错)机制以保障计算的可靠性。实例验证了开源大数据平台Hadoop环境下Delft3D水环境模型集群运算模式,证明该模式的可用性、可扩展性和可靠性。

## 2 水环境模拟集群运算模式与构架

### 2.1 运算模式

基于大数据技术中的分布式并行计算框架实现的集群运算模式属于多算例多任务分解并行计算,即每一个算例对应一个计算任务,这种模式适合大批量模型计

算,应用面更广,其核心是并行分布式存储和计算。位置感知将计算移动到数据所在的位置(存储)是一个重大的进步<sup>[17]</sup>,即通过“数据本地化”可减少数据迁移,以节约网络带宽,获得高效率的计算性能。分布式存储将数据分散存储到多个节点,并且同一份数据在不同节点上保存多个副本,不但实现了数据本地化,还实现了数据冗余备份,同时保障了数据的安全性。分布式计算则通过位置感知将算法/模型等计算资源分发到数据的所在位置,达到“计算本地化优化”的目的。

水环境模拟集群运算模式充分借鉴“数据本地化”和“计算本地化”的思想,将模拟场景模型配置文件分布到各个计算节点以实现“配置文件本地化”,模型计算程序则通过感知定位并将程序分发到模型配置文件所在节点,从而实现“计算程序本地化”。图1描述了水环境模拟集群运算模式。配置文件的分布式存储冗余备份机制缩短了计算程序寻址感知时间。

水环境模型的分布式计算包括位置感知、本地化计算和计算结果分布式存储3个过程。分布式分发机制可以快速定位到

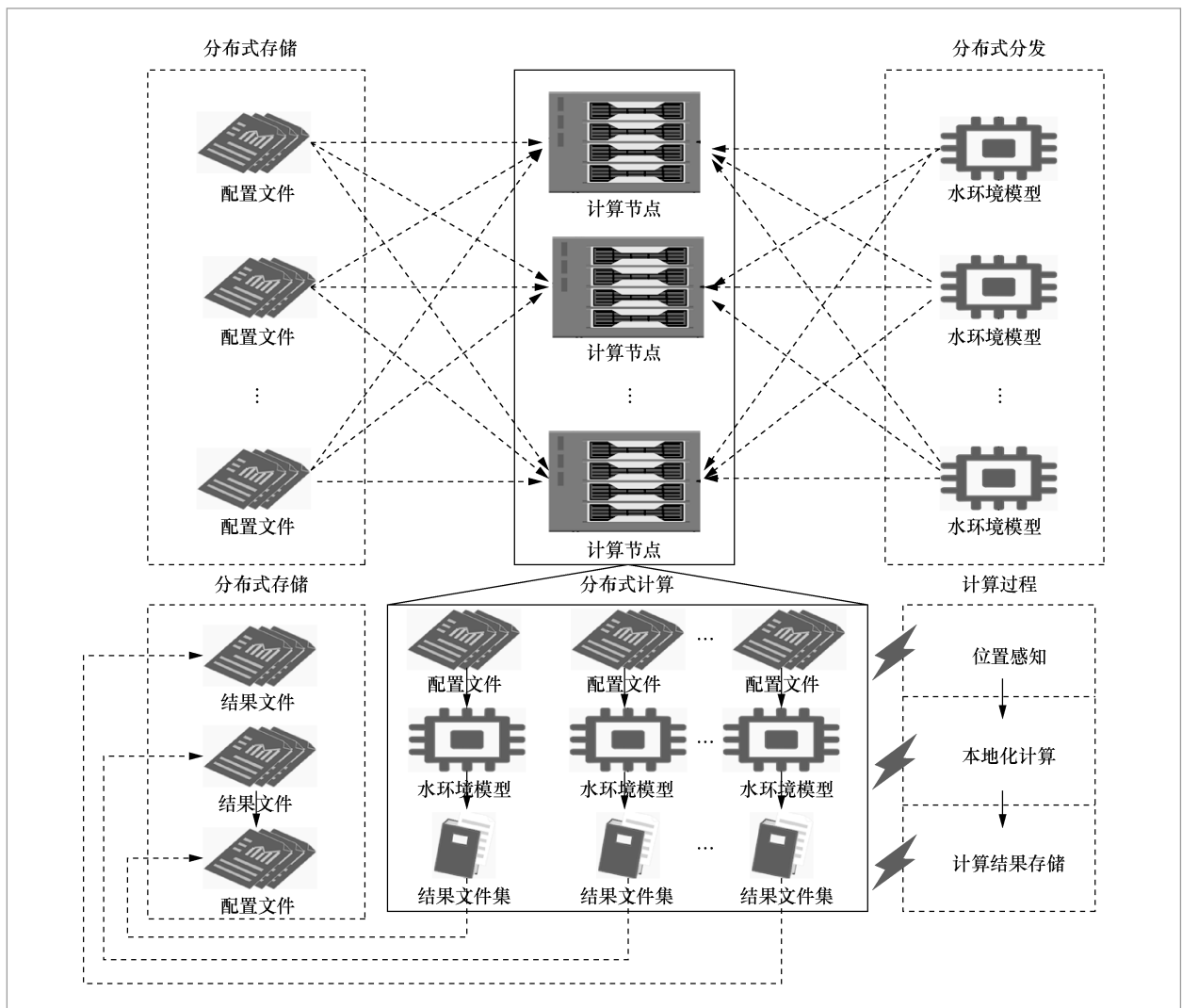


图1 水环境模拟集群运算模式

配置文件所在的计算节点,水环境模型执行文件被自动下载到计算节点,并创建运行空间,启动模型读取配置文件,执行模型本地化计算,最后将计算结果写入分布式存储。

## 2.2 技术架构

综合现有水环境模型和大数据技术的发展现状,图2给出了水环境模拟集群运算模式实现的技术架构。实现流程为:在“数据”层形成“模型/算法”层运行所需的基础配置文件集合;然后文件集合和模型/算法被上传到“存储”层,并以Hadoop分布式文件系统(Hadoop distributed file system, HDFS)格式存储;在“计算”层中,计算作业调度启动离线批量计算程序,

实现多个模型/算法的分布式并发计算,计算结果解析后被存入HBase数据库,“计算”层利用Spark分析HBase数据库中存储的记录,以支持业务应用。

水环境数值模型模拟计算是典型的CPU密集型运算,具有计算性能需求高、运行时间长、计算结果文件大的特点。针对此特点,“计算”层中需要筛选合适的技术框架,其中MapReduce和Spark均属于分布式计算框架,前者处理“本地化”模型输入文件数据,后者处理“本地化”模型输出的结果数据。MapReduce适合离线式批量计算,Spark适合内存式迭代计算,此外还引入了QUARTZ定时任务框架,用于周期性执行的计算作业调度管理。鉴于数值模型计算耗时的特点,采用MapReduce离线式计算框架,而对于计算结果的交互式

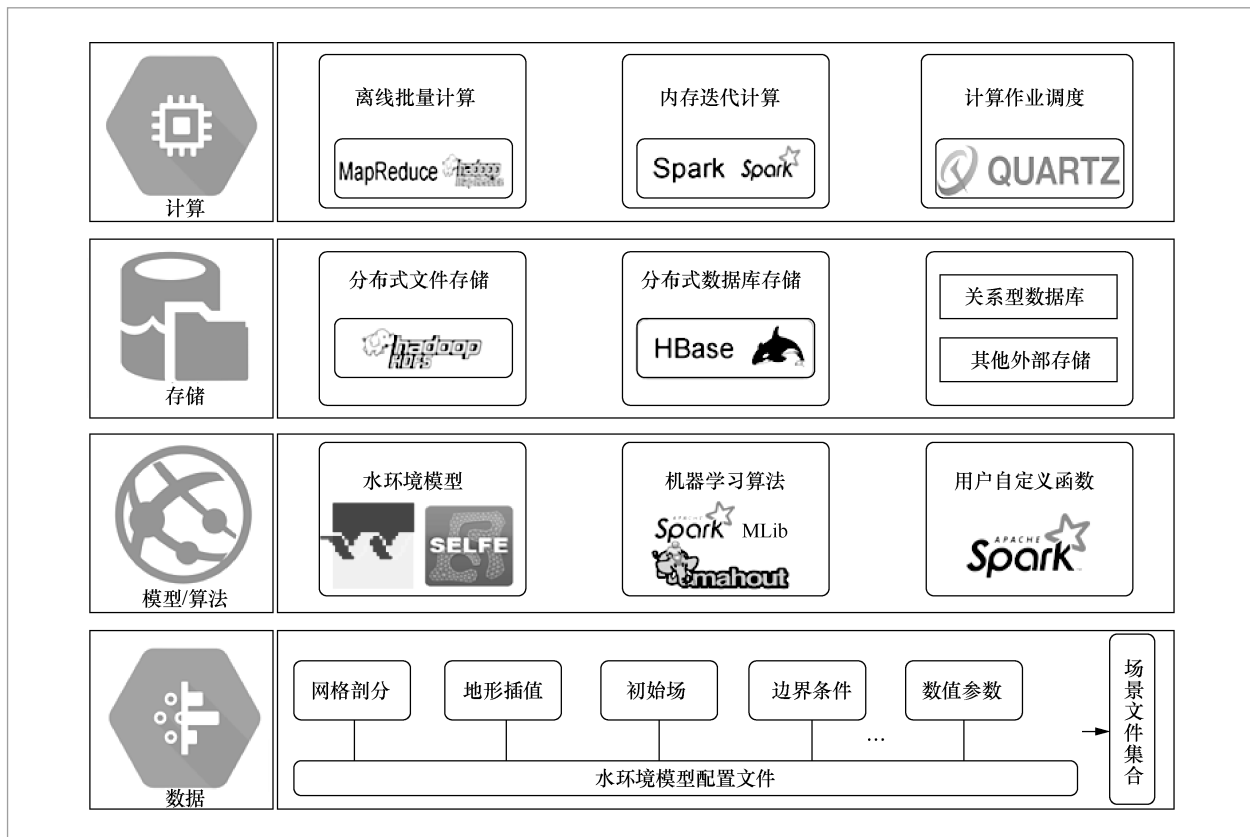


图2 水环境模拟集群运算模式技术架构

分析,则采用Spark内存迭代式计算框架。

在分布式文件存储和数据库存储方面,HDFS和HBase已经得到广泛应用。水环境模型配置和原始模拟结果通常以文件格式存储,适合采用HDFS存储;模拟结果解析后的结构化数据则适合采用HBase数据库进行存储。此外对于其他常用的数据,如模型运算相关基本信息描述、集群软硬件资源描述等,适合采用常规关系型数据库或者其他外部存储方式。

数值模型/算法是水环境模拟的计算引擎和核心价值,一般根据实际应用侧重的功能需求来选择合适的模型。基于开源、具有商业应用背景和完善的社区技术支持等筛选依据,图2推荐Delft3D和SELFE 2种水动力水质模型,此外还包括TELEMAC、EFDC等其他模型。针对大规模集群计算产生的模拟结果分析,在技术架构的“模型/算法”层集成机器学习算法库,如Apache Mahout、Spark MLlib等。为了支持用户自定义算法,在“模型/算法”层也提供用户自定义函数。鉴于模拟结果存储在HBase中,因而推荐采用Spark实现自定义操作。

“数据”层中定义了数值模型/算法运行所需要的一系列基础配置文件。典型的水环境模型配置文件包括研究区网格剖分、地形插值、初始场、边界条件设定等。根据参数率定、不确定性分析及情景决策的应用需要,可形成一系列场景文件集合。

## 3 应用案例

### 3.1 模式设计

以Hadoop MapReduce环境下Delft3D模型运算为例,验证集群运算模式的可行性。Delft3D模型由荷兰WJ|Delft

Hydraulics公司开发,用于河流、湖泊、水库、河口和海岸等自由地表水环境的水动力和水生态计算。Delft3D由7个模块组成,包括:水动力模块(Delft3D-FLOW)、波浪模块(Delft3D-WAVE)、水质模块(Delft3D-WAQ)、颗粒跟踪模块(Delft3D-PART)、生态模块(Delft3D-ECO)、泥沙输移模块(Delft3D-SED)和动力地貌模块(Delft3D-MOR)。这些模块功能独立且相互联系,每一个模块都可单独执行或与一个或多个其他模块组合执行,能够模拟二维和三维水流、波浪、水质、生态、泥沙输移及床底地貌以及各个过程之间的相互作用。Delft3D模型是目前国际上先进的水动力-水质模型之一,在国际上应用十分广泛。图3中水环境模拟集群运算模式设计的核心在于将Delft3D模型的每个参数集文件通过集群分发机制分布到各个数据节点,集群通过位置感知机制将计算程序定位到数据节点,在节点上通过读取参数集文件,重构模型配置文件,并完成模型计算过程。

### 3.2 集群环境

Hadoop采用传统的主/从框架体系,集群搭建中至少采用3台服务器,一台服务器作为集群的主节点,即名称节点(NameNode)。NameNode的失败会导致集群失败。为了保障NameNode故障时恢复集群运转,选择另外一台服务器作为NameNode的备份节点,即第二名称节点(secondary NameNode)。NameNode负责管理整个文件系统,维护和更新文件的分块、存储信息,监视文件系统的健康状态。为了维护整个文件系统,NameNode需要在内存磁盘中进行大量的读写操作,这些操作会抢占计算资源,因此,通常托管NameNode的机器不再

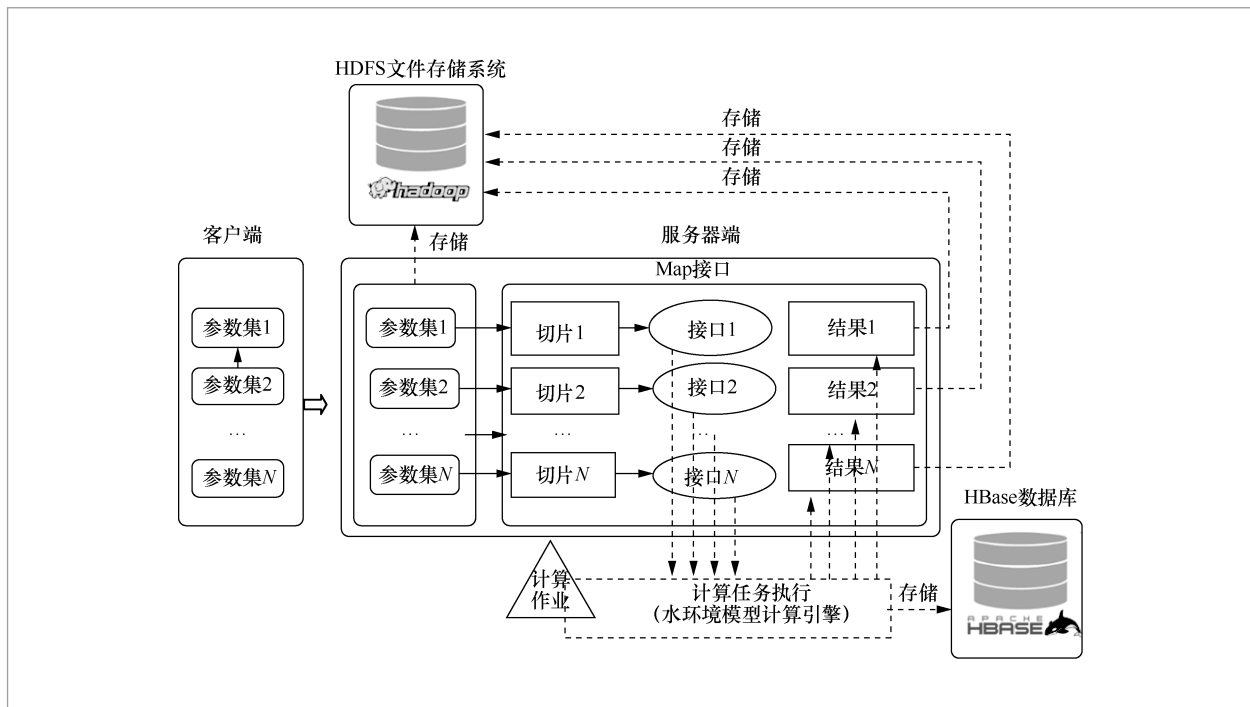


图3 水环境模拟集群 MapReduce 运算核心流程

承担数据节点(DataNode)和任务追踪器(TaskTracker)任务。剩余集群同时作为TaskTracker和DataNode。

表1中采用7台机器搭建Hadoop集群,其中“Master”在集群中承担NameNode和作业追踪器(JobTracker)的角色,“Master2”承担secondary NameNode的角色,“s200~s204”承担DataNode和TaskTracker的角色。7台机器通过千兆交换机进行网络连接,网络结构如图4所示。

表1 集群软硬件环境

机器名	硬件配置	集群中角色
Master	HP ProLiant DL580 g7, 32核, 2.67 GHz, 内存64 GB, 硬盘容量900 GB	NameNode、JobTracker
Master2	HP ProLiant DL580 g7, 32核, 2.67 GHz, 内存64 GB, 硬盘容量900 GB	Secondary NameNode
s200~s204	ThinkCentre M841t, 8核, 3.4 GHz, 内存32 GB, 硬盘容量2 TB	DataNode、TaskTracker

### 3.3 实现流程

图5按照“情景定义-集群计算”2层结构描述了实现流程。“情景定义”负责为模型运行提供一系列不同的场景文件,它包括模型定义、参数提取和参数集定义3个部分。图6中,模型定义过程提供了水环境数值模型运行的一系列基础配置文件,包括研究区网格剖分、地形插值、边界条件设定、初始化状态设定等过程对应的参数设置文件,这些文件由一个项目工程统一进行管理;参数提取过程根据参数率定、不确定性分析及情景决策的需要,从模型定义文件中选取相应的参数,如河床底部糙率、紊动粘滞系数、污染物降解系数等;参数集定义过程将上述参数采用拉丁超立方抽样,生成一系列参数集文件,该文件与基础配置文件压缩后存入HDFS,副本数默认取3;计算任务分发过

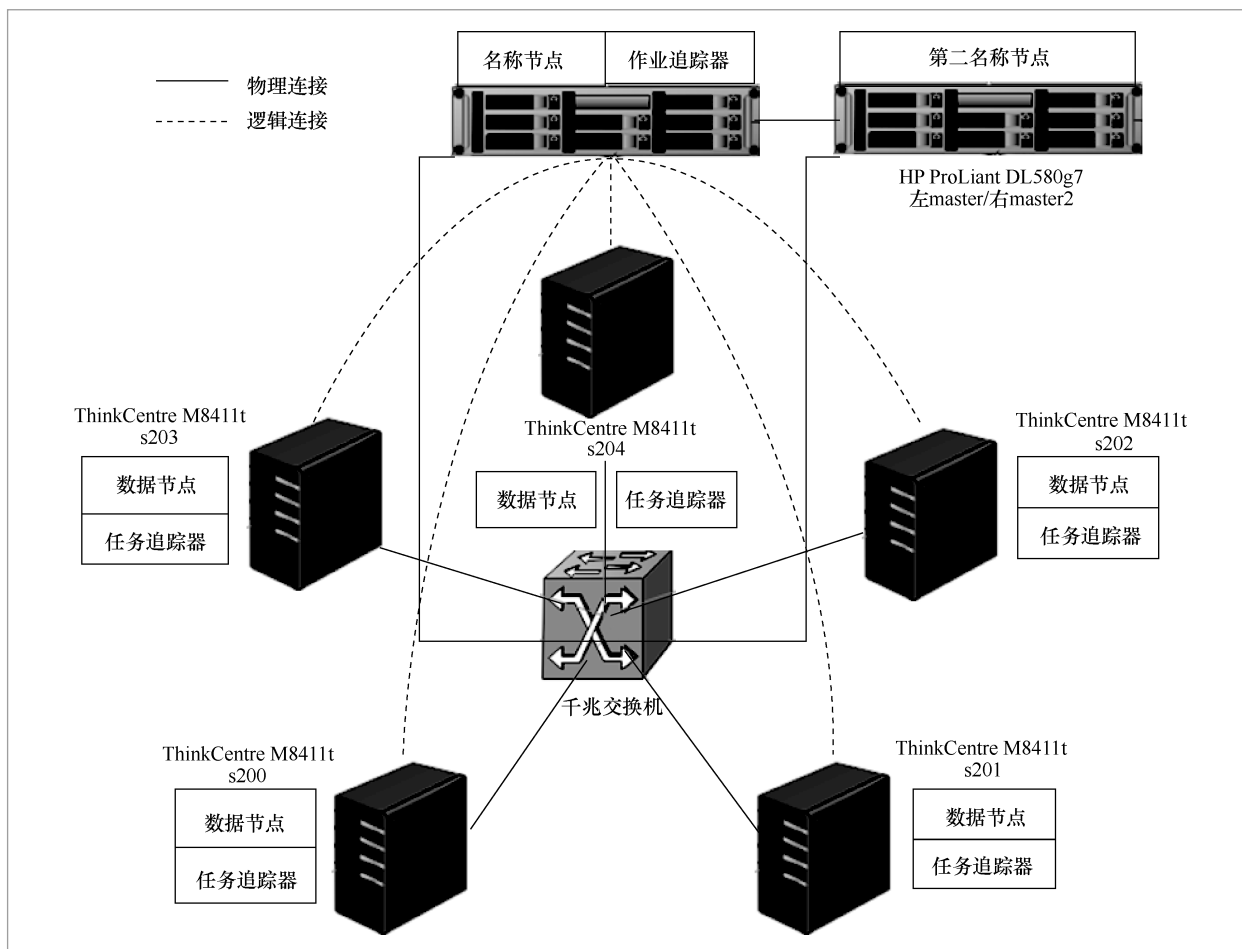


图4 Hadoop 集群网络结构

程中，“集群运算”负责搭建集群运算环境，并将参数集定义过程中的参数集文件分发到集群的计算节点(task slot)；计算任务执行过程在每个计算节点上读取参数集文件，重构模型配置文件，执行模型计算，并将模型计算结果存储到集群。

### 3.4 实验设计

为了测试集群的可用性，同时也为了确保测试过程的可复制性，模型文件采用官方弗里斯兰潮汐口(Frisian tidal inlet)案例；为了测试不同模型大小对集群运算过程的影响，案例的模拟时间分别设计为

1年、半年和1个月；为了测试集群的可靠性和可扩展性，设计了动态增删计算节点测试实验。集群计算性能评价采用加速比指标，加速比是同一个计算任务在单处理器系统和并行处理器系统中运行消耗的时间比率，用来衡量并行系统或程序并行化的性能和效果。加速比计算式为 $S_N = T_1 / T_N$ ，其中 $N$ 指计算节点的数量， $T_1$ 指单计算节点完成计算的用时， $T_N$ 指 $N$ 个计算节点完成计算的用时。

### 3.5 测试结果

图6表明，不论是常规的单作业多任

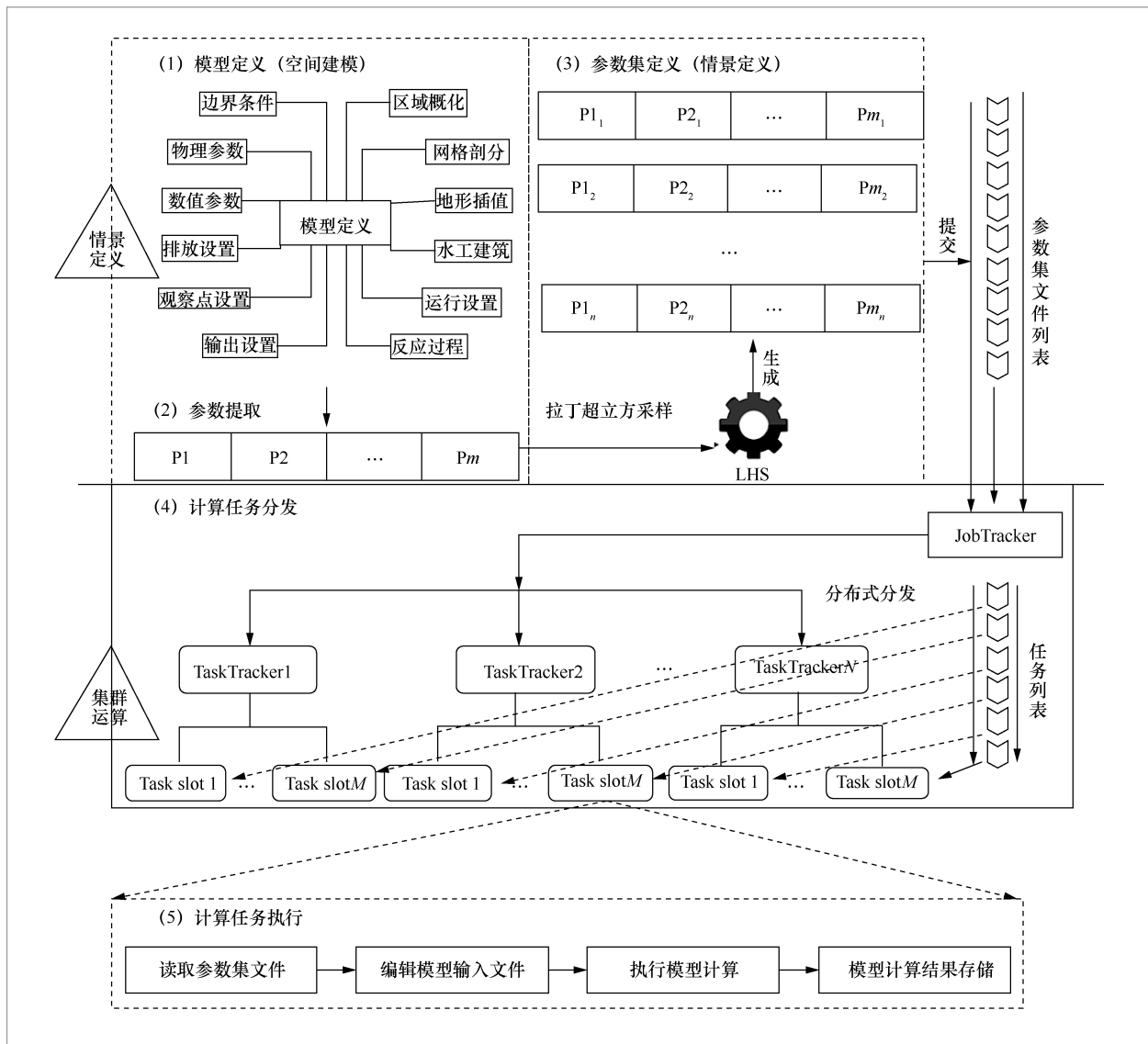


图5 水环境模拟集群运算模式实现流程

务模式（集群只有唯一一个job，内含多个task算例），还是多作业单任务模式（集群拥有多个job，每个job只含有唯一的task算例），集群成功地实现了所有计算任务，证明Hadoop MapReduce环境下Delft3D模型集群运算模式的可行性。此外，上述2种作业模式随着计算节点的增加呈现出以下规律。

首先，随着计算节点的增加，集群的运行时间显著降低。起初，图6(a)的总运行

时间略高于图6(b)的总运行时间，随着计算节点的增加，运行时间的降幅明显高于后者。而且，随着计算节点的增加，模型复杂度和运行时间也呈现出规律：低复杂度模型（模型3）的运行时间降幅明显高于中复杂度模型（模型2），中复杂度模型的运行时间降幅明显高于高复杂度模型（模型1）。总体上，图6(a)中模型的运行时间降幅分别高于图6(b)中相应模型的运行时间降幅。

其次，图7中集群的加速比与计算节

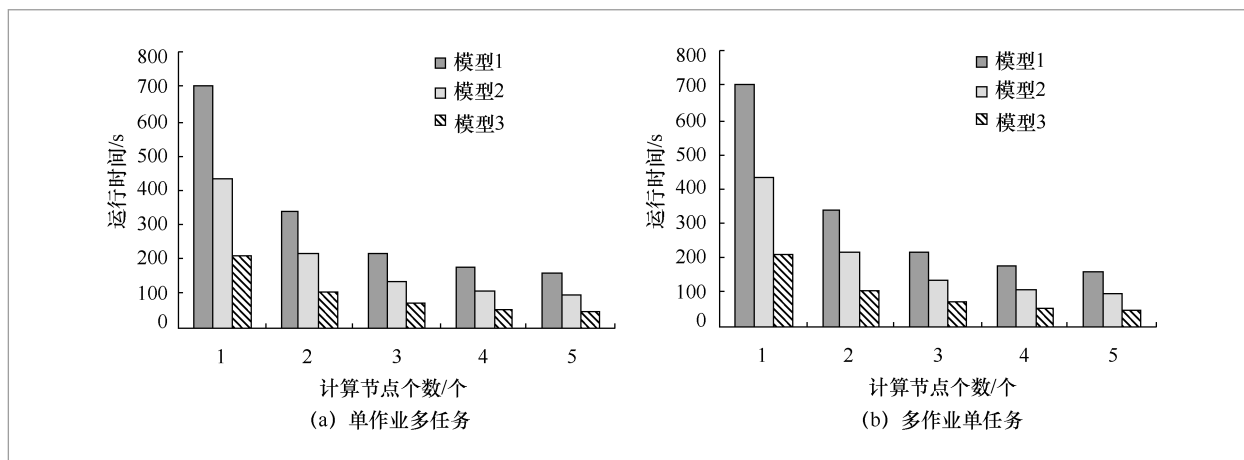


图6 运行时间 / 计算节点个数对比

点呈现近似线性关系：随着节点个数的增加，集群加速比有明显的偏离趋势。相对于图7(b)，图7(a)的加速比更接近理想值。除了计算节点为1时两者加速比均接近理想值1之外，图7(a)明显优于图7(b)。当计算节点为5时，图7(a)3种模型(模型3、模型2、模型1)的加速比分别为4.08、4.22和4.25，图7(b)中的加速比分别为3.89、4.22和3.89。图7(a)稍优于图7(b)，主要归结于前者具有更合理的集群设置。图7(a)中总任务数等于总模拟数，因此作业被分为多个任务，这有利于

任务负载均衡和充分利用计算资源。

而且，模型的大小与加速比在总体上呈现相关关系，集群中低复杂度模型(模型1)获得的加速比明显高于中复杂度模型(模型2)，中复杂度模型获得的加速比明显高于高复杂度模型(模型3)。不考虑模型复杂度，除了当计算节点为1时，2种模式加速比类似接近理想值1之外，随着计算节点增加，3种不同复杂度模型的加速比均与计算节点呈现近似线性关系，且图7(a)中不同复杂度模型获得的加速比均明显高于图7(b)中对应模型获得的加速比。复杂度

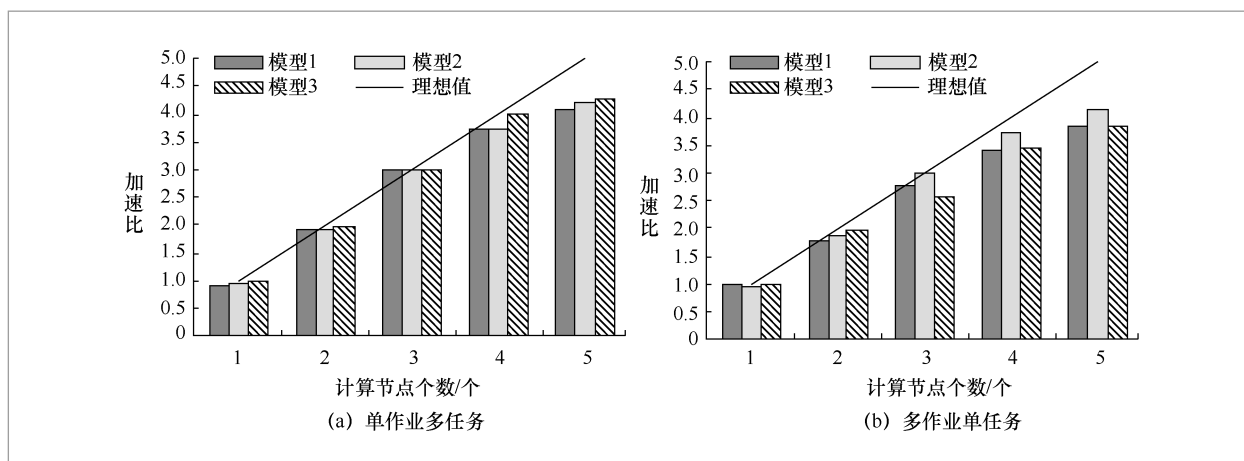


图7 加速比 / 计算节点个数对比

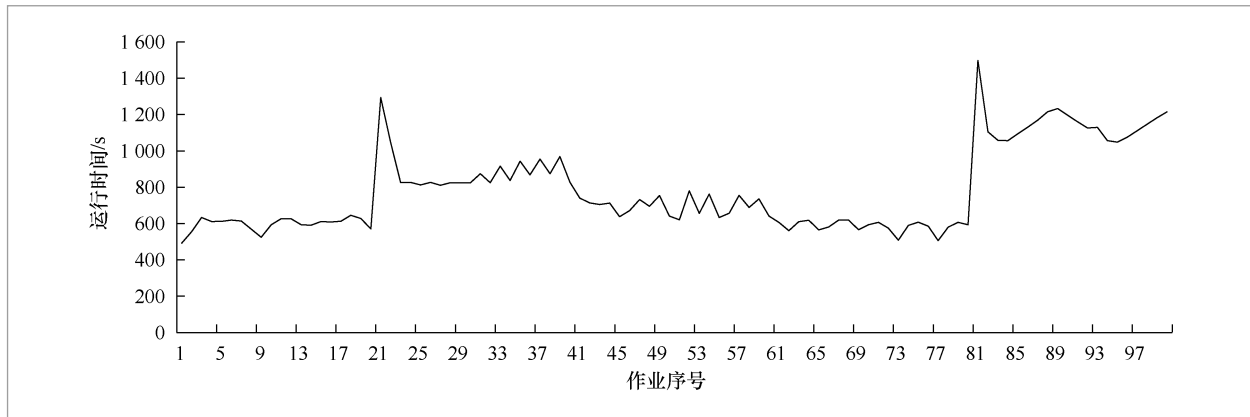


图8 集群中动态增删计算节点测试

越高的模型会损失越多的加速比,其主要原因在于模型计算越复杂,对磁盘I/O需求越大,模拟结果写入磁盘耗时越长。

最后,集群中动态增删计算节点测试(图8)结果显示,在作业1~作业19的运行过程中,每个作业运行时间维持在600 s;当第20个作业运行时,关闭2个计算节点,运行时间迅速升至1 300 s左右,然后平缓降低至900 s左右(位于作业39处);当在作业40处增加一个计算节点时,作业运行时间缓慢降至740 s左右(位于作业59处);当在作业60处再增加一个计算节点时,集群恢复起始状态,作业运行状况和作业1~作业19类似;当在作业79处关闭3个计算节点时,作业80的运行时间迅速升至1 500 s左右,然后缓慢降到1 200 s左右,直至剩余作业计算完成。研究表明,集群运行过程中动态增删计算节点会影响计算作业的执行过程,但是不会造成计算任务的失败和集群环境崩溃,证明集群模式具有优秀的容错性、可扩展性和可靠性。

## 4 结束语

水环境模型结构的日益复杂和应用的不断深入,对高性能计算和海量数据处理

提出了新挑战。积极探索利用大数据技术将各种计算资源聚合,实现协同计算,是现代水环境模拟面临的一个迫切需求。本文基于大数据分布式计算框架,建立了一种适用于水环境模型批量模拟计算的运算模式。实例验证了Hadoop MapReduce环境下Delft3D模型集群运算模式的可行性,结果表明集群运算模式可显著加快运算过程,可提供良好的横向扩展和容错的模型运行环境,是解决水环境模拟大规模批量计算需求的一种理想方法。集群运算模式的软件体系基于开源技术架构,硬件体系基于商业PC或者服务器,因此,集群构建具有过程灵活、节约成本的特点,适用于参数率定、不确定性分析以及情景决策等批量计算应用场景。同时该模式也适用于其他环境建模,具有较强的通用性。

## 参考文献:

- [1] 陈永灿,刘昭伟,朱德军. 水动力及水环境模拟方法与应用[M]. 北京: 科学出版社, 2012. CHEN Y C, LIU Z W, ZHU D J. Hydrodynamic and water environment simulation methods and applications[M]. Beijing: Science Press, 2012.
- [2] 彭泽洲,杨天行,梁秀娟. 水环境数学模型及

- 其应用[M]. 北京: 化学工业出版社, 2007: 98-103.
- PENG Z Z, YANG T X, LIANG X J. Mathematical model of water environment and its application[M]. Beijing: Chemical Industry Press, 2007.
- [3] NGUYEN T T, KEUPERS I, WILLEMS P. Conceptual river water quality model with flexible model structure[J]. *Environmental Modelling & Software*, 2018, 104: 102-117.
- [4] MELCHING C S. Application of a water quality model for determining instream aeration station location and operational rules: a case study[J]. *Water Science and Engineering*, 2018, 11(1): 8-16.
- [5] KEUPERS I, WILLEMS P. Development and testing of a fast conceptual river water quality model[J]. *Water Research*, 2017, 113: 62-71.
- [6] REDER K, ALCAMO J, FLORKE M. A sensitivity and uncertainty analysis of a continental-scale water quality model of pathogen pollution in African rivers[J]. *Ecological Modelling*, 2017, 351: 129-139.
- [7] 贾仰文, 王浩, 严登华. 黑河流域水循环系统的分布式模拟(I)-模型开发与验证[J]. *水利学报*, 2006(5): 534-542.
- JIA Y W, WANG H, YAN D H. Distributed model of hydrological cycle system in Heihe River basin I-model development and verification[J]. *Journal of Hydraulic Engineering*, 2006(5): 534-542.
- [8] 王永桂. 流域大尺度水环境模型的高效能集群计算方法研究及其在三峡库区的应用[D]. 武汉: 武汉大学, 2015.
- WANG Y G. Research on high productivity cluster computing technology of large scale water environment model and the application in the Three Gorges Area[D]. Wuhan: Wuhan University, 2015.
- [9] WANG H, FU X D, WANG G Q, et al. A common parallel computing framework for modeling hydrological processes of river basins[J]. *Parallel Computing*, 2011, 37(6): 302-315.
- [10] 左一鸣, 崔广柏. 二维水动力模型的并行计算研究[J]. *水科学进展*, 2008, 19(6): 846-850.
- ZUO Y M, CUI G B. Parallel computation for 2D flow model[J]. *Advances in Water Science*, 2008, 19(6): 846-850.
- [11] 许栋, PAYET D, 及春宁, 等. 浅水方程大规模并行计算模拟城市洪水演进[J]. *天津大学学报(自然科学与工程技术版)*, 2016, 49(4): 341-348.
- XU D, PAYET D, JI C N, et al. Simulation of flood propagation in cities using large scale parallel computation of shallow water equations[J]. *Journal of Tianjin University*, 2016, 49(4): 341-348.
- [12] 靖争. 湖泊水环境数值模拟及非静压水流并行计算方法研究[D]. 武汉: 华中科技大学, 2017.
- JING Z. Numerical simulations of the lake water environment and parallel computations of the non-hydrostatic flows[D]. Wuhan: Huazhong University of Science and Technology, 2017.
- [13] 赵旭东, 梁书秀, 孙昭晨, 等. 基于GPU并行算法的水动力数学模型建立及其效率分析[J]. *大连理工大学学报*, 2014, 54(2): 204-209.
- ZHAO X D, LIANG S X, SUN Z C, et al. Foundation and analysis of computational efficiency for hydrodynamic model based on GPU parallel algorithm[J]. *Journal of Dalian University of Technology*, 2014, 54(2): 204-209.
- [14] 宋利祥, 范光伟, 张文明, 等. 洪水演进高速模拟系统设计与实现[J]. *水电能源科学*, 2017, 35(12): 33-36.
- SONG L X, FAN G W, ZHANG W M, et al. Design and implementation of high speed simulation for flood routing[J]. *Water Resources and Power*, 2017, 35(12): 33-36.
- [15] ZHANG D, CHEN X, YAO H, et al. Moving SWAT model calibration and uncertainty analysis to an enterprise Hadoop-based cloud[J]. *Environmental Modelling & Software*, 2016, 84(C): 140-148.
- [16] WHITE T. Hadoop: the definitive guide[M]. Sebastopol: O'Reilly Media, Inc., 2010.
- [17] MURTHY A, VAVILAPALLI V, EADLINE D, et al. Apache Hadoop YARN[M]. Boston: Addison-Wesley, 2014.

## 作者简介



马金锋(1978- ),男,中国科学院饮用水科学与技术重点实验室(中国科学院生态环境研究中心)助理研究员,主要研究方向为水环境数值模拟。



唐力(1967- ),男,深圳市环境监测中心站高级工程师,主要研究方向为环境监测。



饶凯锋(1976- ),男,中国科学院饮用水科学与技术重点实验室(中国科学院生态环境研究中心)助理研究员,主要研究方向为水生态毒理学、环境预警监测与物联网。



洪纲(1974- ),男,石家庄市环境信息中心高级工程师,主要研究方向为环境大数据、环境监测与管理。



马梅(1967- ),女,中国科学院饮用水科学与技术重点实验室(中国科学院生态环境研究中心)研究员,主要研究方向为水生态毒理学。

收稿日期: 2019-08-06

基金项目: 国家自然科学基金资助项目(No. 51209194); 中国科学院前沿科学重点研究计划项目(No. QYZDY-SSW-DQC004); 广东省省级科技计划基金资助项目(No. 2016B02024007)

**Foundation Items:** The National Natural Science Foundation of China (No. 51209194), Frontier Science Key Program of the Chinese Academy of Sciences (No. QYZDY-SSW-DQC004), Science and Technology Planning Project of Guangdong Province (No. 2016B02024007)