

基于 Hadoop 的校园云计算系统^①

曹风兵, 吴开贵, 吴长泽

(重庆大学 计算机学院, 重庆 400044)

摘要: 针对传统的海量数据处理方法硬件成本太高, 并行程序编写困难的缺点, 在云计算理论的基础上设计了一个用于处理海量数据的校园云计算系统。此云计算系统是在 Hadoop 分布式计算框架的基础上采用 Map-Reduce 编程模型实现对海量数据的并行处理, 有效解决了成本问题, 降低了并行编程的难度。

关键词: 云计算; 分布式计算; 海量数据; Hadoop; Map-Reduce

Cloud System for Campus Based on Hadoop

CAO Feng-Bing, WU Kai-Gui, WU Chang-Ze

(College of Computer Science, Chongqing University, Chongqing 400044, China)

Abstract: As the traditional method of massive data processing has shortcomings of high cost in hardware and the difficulties in parallel programming, a campus cloud computing system platform to handle massive data is designed based on the theory of cloud computing. This cloud computing system is based on the Hadoop distributed computing framework, using map-reduce programming model achieve parallel processing of the massive data. This system can save cost and reduce the difficulty of parallel programming.

Keywords: cloud computing; distributed computing; massive data; Hadoop; map-reduce

1 引言

海量数据是指规模巨大的数据集, 它包括各种空间数据, 统计数据, 文字, 声音, 图像和其他的文本数据。随着信息技术的不断发展, 各种信息量迅速增长, 对这些海量数据进行处理和分析显得越来越重要。

高校作为科研的重要机构, 承担着重大的科研责任, 但是每个学科在科研的过程中总会遇到处理海量数据的问题, 这些海量信息的处理都需要超强的计算能力。如果采用传统的并行处理方式, 面对日益增大的数据量, 存在如下问题:

(1) 编写并行处理程序比较困难。传统并行处理方式, 例如使用 MPI, 由于它的编程模型复杂, 人们会把大量的时间和精力用在编写并行程序上, 并且 MPI 有着通信延迟大, 负载不均衡和程序的可靠性差等缺点。

(2) 硬件的成本太高。由于海量数据的数据量非常大, 采用传统的并行处理就需要高性能计算机, 这

样成本就非常的高。

云计算技术的发展对解决海量数据问题有很大的帮助。本文在云计算理论的基础上提出了一种用于处理海量数据的校园云计算系统模型, 这种云计算系统是基于 Hadoop 分布式计算框架的。Hadoop 是一个云计算平台, 规模可以灵活扩展且对硬件配置要求不高, 成本较低。系统采用 MapReduce 编程模型实现对海量数据的并行处理, 处理程序编写简单, 节省了编写并行程序的时间。本文包括以下几个部分, 第一部分是相关技术介绍, 第二部分是系统分析与设计, 第三部分是系统实现, 最后是总结与展望。

2 相关技术介绍

2.1 Hadoop 计算框架

Hadoop^[1]是一个分布式系统基础架构, 由 Apache 基金会开发。用户可以在不了解分布式底层细节的情况下, 开发分布式程序。充分利用集群的威力进行高

^① 基金项目:国家自然科学基金(90818028)

收稿时间:2010-10-13;收到修改稿时间:2010-11-12

速运算和存储。简单得地说, Hadoop 是一个更容易开发和处理大规模数据的软件平台。云计算概念的出现, 使得 Hadoop 成为了一个面向分布式的云计算平台。

Hadoop 包含两个组件, 分别是分布式文件系统 (Hadoop Distributed File System), 简称 HDFS, 和一个 MapReduce 并行编程模型。分布式文件系统允许你在多台普通机器上存储和复制文件。MapReduce 可以让你很容易就执行并行程序任务。

分布式文件系统 (HDFS)^[2]采用的是主/从架构, 包含一个名字节点 (NameNode) 和多个数据节点 (DataNode)。NameNode 是管理节点, 用于管理数据节点和客户端对文件的访问。当客户端想访问一个在文件系统上的文件时, 它首先和名字节点联系, 名字节点将告诉它在那个数据节点上可以取得文件。名字节点负责调度和保存那些数据块保存和复制到在那些机器上。数据节点是数据仓库, 保存真实的文件数据块。HDFS 有着高容错性的特点, 并且设计用来部署在低廉的硬件上。而且它提供高传输率来访问应用程序的数据, 适合处理那些超大规模数据集。

Hadoop 的特点^[3]有以下几个:

(1) 扩容能力: 能可靠地存储和处理大规模数据集。

(2) 成本低: 可以通过普通机器组成的服务器群来分发以及处理数据。

(3) 高效率: 通过分发数据, Hadoop 可以在数据所在的节点上并行地处理它们, 这使得处理非常的快速。

(4) 可靠性: Hadoop 能自动地维护数据的多份复制, 并且在任务失败后能自动地重新部署计算任务。

2.2 MapReduce 编程模型

MapReduce^[4]是 Google 提出的一种编程模型, 用于大规模数据集的并行运算。它包括两个阶段: Map 阶段和 Reduce 阶段。图 1 展示了 MapReduce 的执行过程, MapReduce 的计算过程就是先将大规模数据集分解成多个小数据集, 然后把这些数据集分发给多个

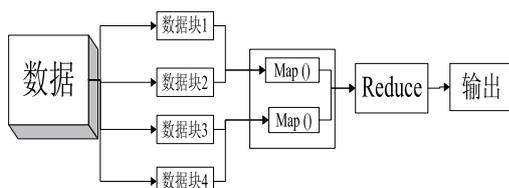


图 1 MapReduce 执行过程

map 节点进行并行处理产生中间结果, 最后在 reduce 阶段对这些中间结果汇总, 得到最终结果。

2.3 云计算

云计算^[5]虽然是近几年被炒的比较热的技术名词, 但是从本质上讲, 云计算并不是一个全新的技术。它是在并行计算, 网格计算, 分布式计算以及虚拟化基础上发展起来的, 也可以说是这些概念的一个新的发展模式。这种模式是将很多位于同一地点或者不同地点的空闲计算资源集合起来通过网络为用户或者企业提供产品, 服务和解决方案。

云计算通过一个强大的基础架构平台提供云计算服务, 其基本形式包括^[6], 基础设施即服务 (IAAS)、平台即服务 (PAAS) 和软件即服务 (SAAS)。云的应用有很多, 主要有: 效用计算、服务商业平台、云存储以及在网络安全方面的病毒查杀等等。

3 系统分析与设计

3.1 系统开发目标与原则

由于海量数据处理需要超强的计算能力, 如果采用传统的并行处理方式存在并行程序编写困难, 用户大部分精力用在了编写程序方面, 操作起来很不方便。云计算的特点就是把计算资源整合起来从而提供高性能计算。用于处理海量数据的校园云计算系统的研发目标就是能够实现资源的充分利用, 降低编写处理程序的难度, 高效地处理数据, 方便用户使用和操作, 便于数据管理, 及时对用户处理结果。

系统设计的原则有以下几点: 第一, 经济原则。充分利用现有的资源来构建系统基础设施, 由于采用 Hadoop 作为底层集群部署, 对硬件的要求不高, 一般的 PC 机就可以达到要求。第二, 高效原则。系统以云计算为基础, 充分利用云计算平台的特性和现有的资源, 采用合理的体系结构, 提高海量数据的处理效率。

3.2 系统总体设计

校园云计算系统模型如图 2 所示, 本系统采用三层体系架构, 包括客户端, 系统管理层和底层基础设施服务层。

第一层, 客户端, 它是通过 Web 浏览器向系统服

务器发送请求，服务器对用户进行身份验证后通过 HTTP 协议把所需的信息返回给客户端，客户机接受信息，将它显示在浏览器上。

第二层，系统管理层，具有应用程序扩展能力的 Web 服务器，它通过执行相应的应用程序与底层集群进行连接，然后集群负责处理用户提交的数据并把结果提交给 Web 服务器，再由服务器传回客户端。

第三层，底层基础设施层，Hadoop 集群作为系统的底层基础设施，为系统提供计算和存储。Hadoop 集群中的分布式文件系统 HDFS(Hadoop Distributed File System)，用于管理数据和分配任务。HDFS 中的名字节点 (NameNode) 作为管理节点，其他的数据节点 (DataNode) 作为存储节点和计算节点。

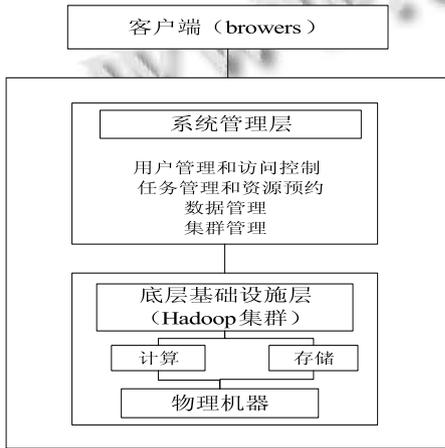


图 2 系统架构模型

3.3 功能子模块设计

整个系统管理层包括四大功能模块：

- (1) 用户管理模块：包括用户注册，身份认证，用户信息管理，权限管理，交互控制。
- (2) 任务管理模块：包括用户任务申请，资源申请，信息反馈，返回处理结果。
- (3) 数据管理模块：包括数据上传，数据下载，和删除操作。
- (4) 集群管理模块：包括动态显示集群工作状态和任务进度。

系统功能如图 3 所示，采用树形结构图形描述了整个系统管理层的功能模块。

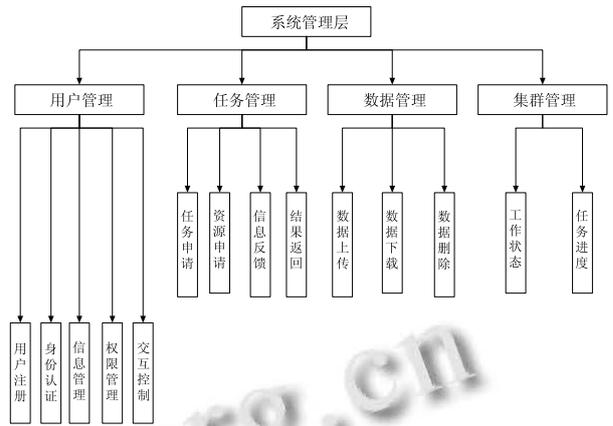


图 3 系统功能模块图

4 系统实现

4.1 底层 Hadoop 集群部署^[7]

本系统采用的是 Hadoop-0.20.0 版本，使用 5 台普通电脑，操作系统都为 Ubuntu 8.04,这 5 台机器的机器名和 IP 地址和角色如下：

- Hadoop1 172.20.53.241 master; namenode
- Hadoop2 172.20.53.242 slave; datanode
- Hadoop3 172.20.53.243 slave; datanode
- Hadoop4 172.20.53.244 slave; datanode
- Hadoop5 172.20.53.245 slave; datanode

其中 Hadoop1 作为集群的主节点 (名字节点: NameNode)，其余机器作为从节点 (数据节点: DataNode)。

集群部署步骤如下：

- 第一步，需要在每台机器上创建同样的目录结构。
- 第二步，在每台计算机上配置 Java 环境。本系统采用的 JDK1.6 版本。

第三步，在主节点上将下载的 hadoop-0.20.0 压缩包解压到/home/hadoop 目录下。

第四步，为集群配置 SSH，实现机器之间的免密码登陆。在主节点的操作命令为：

```

$sudo apt-get install ssh
$ssh-keygen-t rsa-p
$cat/home/hadoop/.ssh/id_rsa.pub>>~hadoop/.ssh/authorized_keys

```

完成后，只需在主节点上运行 ssh-keygen 程序，

将在主节点上创建的 keys 通过 scp 命令拷贝到从节点的同样目录上, 这样从主节点到从节点的登陆就不需要密码了。

第五步, 修改 `hadoop-0.20.0/conf/hadoop_env.sh` 文件, 在其中设置 `JAVA_HOME` 环境变量和 `HADOOP_HOME` 变量:

```
export JAVA_HOME=/usr/lib/jvm/java-6-sun/jre
export
HADOOP_HOME=/home/hadoop/hadoop-0.20.0
```

第六步, 添加主节点 (NameNode), 修改 `hadoop-0.20.0/conf/masters` 文件, 添加内容为: 172.20.53.241

第七步, 添加从节点 (DataNode), 修改 `hadoop-0.20.0/conf/slaves` 文件, 把各从节点的 IP 地址添加进去, 每行一个。

第八步, 配置 `hadoop-0.20.0` 目录下 `conf/core-site.xml`, `conf/hdfs-site.xml`, `conf/mapred-site.xml` 文件。

第九步, 将在主节点配置好的 `hadoop-0.20.0` 用 scp 命令发送到各从节点, 命令如下:

```
scp -r /home/hadoop/hadoop-0.20.0
hadoop2@172.20.53.242:/home/hadoop/
scp -r /home/hadoop/hadoop-0.20.0
hadoop3@172.20.53.243:/home/hadoop/
scp -r /home/hadoop/hadoop-0.20.0
hadoop4@172.20.53.244:/home/hadoop/
scp -r /home/hadoop/hadoop-0.20.0
hadoop5@172.20.53.245:/home/hadoop/
```

第十一步, 完成以上工作后, 在主节点上启动整个集群, 检测集群是否可以正确运行:

```
$cd /home/hadoop-0.20.0/bin 回车
```

运行命令:

```
$hadoop namenode -format //格式化分布式文件系统
```

```
$bin start-all.sh //启动 Hadoop 守护进程有 //namenode,datanode.jobtracker,tasktracker
```

4.2 功能模块实现

校园云计算系统是建立在 Hadoop 平台上的一个

应用系统, 使用 WEB 技术开发。系统提供一个云计算应用服务平台, 通过网络为用户提供数据处理服务。

本系统是采用的 JSP/serverlet 技术进行系统功能模块的开发, 采用 Tomcat 作为服务器, 用 MySQL 作为数据库来存储用户信息及其他信息。系统运行界面如图 4 所示:



图 4 系统运行界面

需要实现的功能模块有以下四个: 用户管理模块, 任务管理模块, 数据管理模块, 集群管理模块。

(1) 用户管理模块

系统为新用户提供注册页面, 如图 5 所示, 用户将个人信息填入, 系统将把这些信息保存在数据库中, 完成注册后, 获得一个用户名。用户在登陆系统时, 系统会根据数据库中的用户信息对用户进行身份认证。只要认证通过才能使用系统其他功能。



图 5 用户注册页面

(2) 任务管理模块

如果用户需要让系统完成某些计算任务, 就需要提交一份任务申请, 系统管理委员会根据实际情况对用户提交的任务进行审核, 只要审核通过才会给用户分配计算资源。在任务完成后系统会将结果返回给用户。

户。

(3) 数据管理模块

这个模块主要实现三个功能：数据上传，数据下载和删除数据。这些操作都可以通过调用 Hadoop 的 API 接口来实现。实现核心代码如下：

数据上传：

```
Public boolean FileUpload (String localFile, String
hadoopFile)
```

//将本地文件传到 HDFS 中

{.....}

```
Configuration conf=new Configuration ();
```

```
FileSystem src=FileSystem.getLocal (conf);
```

//本地文件

```
FileSystem dst= FileSystem.get (conf);
```

```
Path srcpath = new Path (localFile);
```

//本地文件路径

```
Path dstpath = new Path (HDFSFile); //hadoop 路
```

径

```
FileUtil.copy (src, srcpath, dst, false, conf);
```

//拷贝文件到 HDFS??

.....

数据下载代码：

```
public class getdata {
```

.....

```
String uri=arg [0];
```

```
Configuration conf=new configuration ();
```

```
FileSystem fs =FileSystem.get (UAI.create (uri),
conf);
```

//使用 FileSystem API 来获得文件数据

```
InputStream in=null;
```

.....

```
in =fs.open (new path (uri));
```

```
IOUtils.copyBytes (in, System.out, 40961, false);
```

.....

文件删除代码：

.....

```
Configuration config =new Configuration ();
```

```
FileSystem fs=FileSystem.get (config);
```

```
Path path=new Path(filename); //文件路径
```

```
boolean delete=fs.delete (path, true);
```

.....

(4) 集群管理模块

该模块主要是在任务执行过程中显示集群的工作状态和任务进度。通过调用 Hadoop 内置的 WEB 接口就能实现。NameNode 和 DataNode 分别有一个内置的 WEB 服务器，来展示集群当前状态的一些基本信息。访问 http://172.20.53.241:50030 可以查看 map/reduce 的管理；访问 http://172.20.53.241:50060 可以查看 map 任务和 reduce 任务的运行状态；访问 http://172.20.53.241:50070，可以查看主节点以及整个分布式文件系统的状态，也可以浏览文件系统中的文件以及日志等。

4.3 系统执行过程

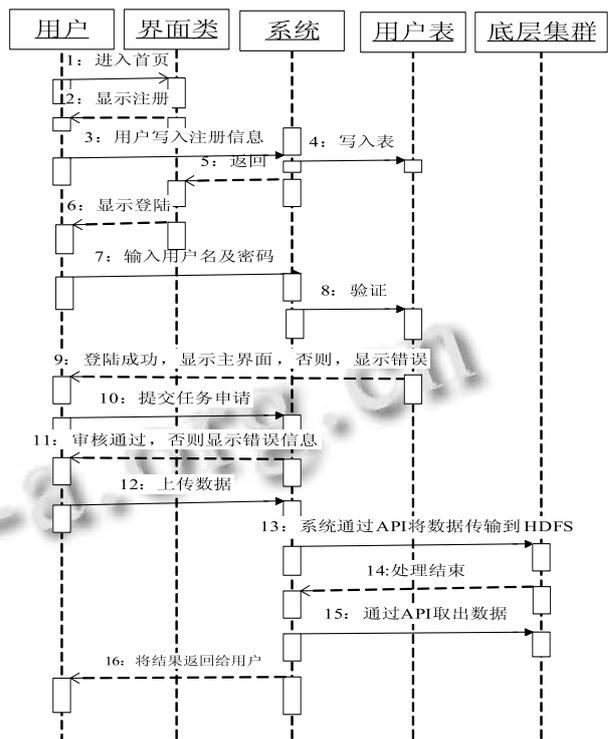


图 6 系统执行时序图

图 6 展示的是系统执行时序图。系统业务执行的逻辑过程如下：

(1) 用户通过浏览器进入系统主界面，如果是新用户，需要完成注册，然后登陆。用户在登陆时，系统会根据用户账号对用户进行身份认证，认证通过后

用户才能使用系统其他功能模块。

(2) 用户如果需要完成某些任务活计算, 就需要向系统提出任务申请, 在系统提供的任务申请表中详细填写任务内容, 由于 Hadoop 并不是对所有的海量数据都适合, 有的数据因为数据自身的结构并不适合使用 Hadoop 进行处理, 所以系统管理员会对用户提出的任务申请进行审核, 只要审核通过后才能使用系统计算资源。

(3) 用户的任务通过审核后, 用户需要将需要计算的数据通过系统服务器传输到 Hadoop 的分布式文件系统(HDFS)中。

(4) Hadoop 集群中的分布式文件系统 (HDFS) 会对数据进行分割并分发到各个数据节点, 然后启动 Map 任务对数据进行处理, 在处理过程中用户可查看任务的处理进度以及集群各节点的工作状态。处理完成后启动 Reduce 任务将结果汇总, 最后将汇总后的结果通过系统服务器返回给用户。

4.4 系统运行效果分析

(1) 系统管理方面

能够保证系统的安全, 各功能模块均能正常运行。方便校内用户的使用, 对学校各学科的科研工作提供了方便。

(2) 资源利用和成本方面

由于 Hadoop 集群具有很灵活的规模伸缩能力, 能够很方便将校园内闲散的计算资源整合起来, 使资源得到充分的利用。Hadoop 对硬件的要求不高, 一般的 PC 机就能满足要求, 所以节约了成本。

(3) 集群性能方面

为测试集群系统的性能, 以气象日志数据^[8]作为实验数据。这些数据是从国家气候中心 (National Climatic Data Center) 下载的网址为: <http://www.ncdc.noaa.gov/>。这个网站的气象数据是由分布在全球各地的气象传感器收集的, 所以数据量非常的大, 并且它们都是半结构化且面向记录的数据, 使用 Hadoop 和 MapReduce 进行分析非常合适。根据本系统的机器配置, 通过使用不同数量级的数据对系统进行测试并和单机在性能方面进行比较。为了取得更准确的结果, 采用多次实验, 取平均值进行比较。结果如图 7 所示:

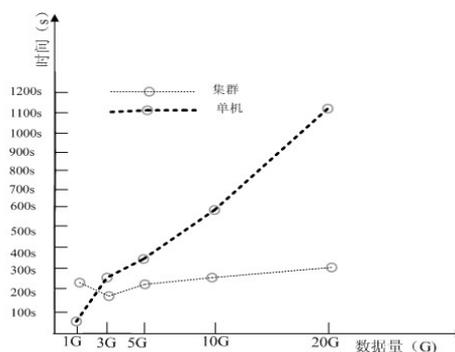


图 7 集群性能测试图

由图 7 可以看出:

(1) 当数据量比较少时, Hadoop 集群所用时间比数据量较大时还要多, 并且速度还不如单机。这是因为 Hadoop 是具有分布式特性的, 适合处理大规模数据集, 如果数据量太少, Hadoop 就不能发挥自身优势。

(2) 当数据量增大时, Hadoop 的性能开始发挥出来, 从图 7 可以看出, 在数据量从 5G 增加到 20G 时, 集群处理时间并没有太大的变化, 但是单机的处理时间增长幅度明显增大, 这时集群的优越性开始显现出来; 在数据达到 20G 时集群的速度大约是单机的 4 倍。由于条件有限, 没能使用更大的数据量对系统进行测试, 20G 对 Hadoop 集群来说还是一个小规模数据集, 但从现阶段的对比实验中已经可以看出 Hadoop 在海量数据处理中的性能特点和优势。

5 总结与展望

通过对云计算基础知识、Hadoop 云计算平台以及 MapReduce 编程模型的学习和研究, 设计和实现了一个用于处理海量数据的校园云计算系统。本系统采用三层体系架构, 有着规模可灵活扩展、开发成本低、操作方便、降低了并行程序编写的难度以及能够高效地对海量数据进行分析 and 处理的优点。

目前, 系统的功能比较单一, 还需要进一步完善。由于 Hadoop 中计算和存储的紧耦合的缺点, 还有许多问题需要解决。例如, 如何在系统中通过部署一个超级数据节点 (SuperDataNode)^[9]来实现 Hadoop 中计算和存储的分离, 这是一个等待解决的问题。下一步

(下转第 5 页)

台级应用服务集成 Web 服务, 主要包含有代理服务、业务基础服务、文件服务、 workflow 服务、报表服务、消息服务、任务调度服务等, 同时平台针对这些应用服务都提供了统一的基于 Web 页面的监控管理界面。此外, 平台本身的实现是数据库无关的, 即可以在不更改源代码的情况下运行于不同数据库之上。

5 结束语

OMS 的构建和应用是一个长期而艰巨的过程, 必须有计划、有步骤的按阶完成。它不仅是管理信息系统的建设过程, 也是管理工作流程的规范化和标准化的过程。本文分析了辽宁电网 OMS 的设计和关键技术, 为今后大型电网调度管理系统一体化建设提供参考。

参考文献

1 国家电力调度中心. 智能电网调度技术支持系统建设框架.

2008.

- 2 谢开, 刘永奇, 朱治中, 等. 面向未来的智能电网. 中国电力, 2008, 41(6): 19-22.
- 3 陈树勇, 宋书芳, 李兰欣, 等. 智能电网技术综述. 电网技术, 2009, 33(8): 1-8.
- 4 胡学浩. 智能电网——未来电网的发展态势. 电网技术, 2009, 33(14): 1-5.
- 5 吴琼, 刘文颖, 杨以涵. 智能型电网调度决策支持系统的开发与实现. 电力系统自动化, 2006, 30(12): 79-83.
- 6 何仰赞, 温增银. 电力系统分析. 第 3 版. 武汉: 武汉华中科技大学出版社, 2002.
- 7 李向荣, 郝悍勇, 樊涛, 等. 构筑数字化电网建设信息化企业. 电力系统自动化, 2007, 31(17): 125.
- 8 姚建国, 杨胜春, 高宗和, 等. 电网调度自动化系统发展趋势展望. 电力系统自动化, 2007, 31(13): 7-11.
- 9 李永锋. 对搞好地调一级调度管理的探讨. 电力安全技术, 2003, 5(3): 21-22.

(上接第 11 页)

的主要工作就是完善系统功能; 在系统架构中引入虚拟化技术; 研究如何解决 Hadoop 中存储和计算的紧耦合; 实现底层集群能够按照用户的需求自动部署, 使系统尽可能符合云计算概念。

参考文献

- 1 Hadoop 官方网站. 2010. <http://hadoop.apache.org>
- 2 HDFS Architecture. 2010. http://hadoop.apache.org/common/docs/current/hdfs_design.html
- 3 Borthakur D. The Hadoop Distributed File System: Architecture and Design. Apache Software Foundation. 2007, 3-14.
- 4 Dean J, Ghemawat S. Mapreduce: Simplified Data Processing on Large Clusters. Communications of ACM,

2008, 51(1): 107-113.

- 5 陈全, 邓倩妮. 云计算及其关键技术. 计算机应用, 2009, 29(9): 2653-2654.
- 6 Ajkumarbuyya SP, Vecchiola C. Cloudbus Toolkit for Market-Oriented Cloud Computing. Lecture Notes in Computer Science, 2009, 5931(2009): 24-44.
- 7 Hadoop Cluster Setup. 2010. http://hadoop.apache.org/common/docs/current/cluster_setup
- 8 White T. Hadoop: The Definitive Guide. 北京: 清华大学出版社, 2010. 15-20.
- 9 Porter G, UC San Diego, La Jolla. Decoupling Storage and Computation in Hadoop with SuperDataNodes. ACM SIGOPS Operating System Review, 2010, 44(2): 41-46.