基于 Hadoop 的海量网格数据建模①

胡志刚 梁晓扬 (中南大学 信息科学与工程学院 湖南 长沙 410083)

摘 要: 针对网格实验的实际需要和现有网格仿真工具存在的不足，提出了一种结合 Hadoop 技术进行海量网格数据建模的方法。利用提出的建模方法，研究人员可以从海量数据中挖掘出实验所需核心数据，并建立这些数据所满足的数学模型。在网格仿真实验中使用这些数学模型生成网格负载，将会提高网格仿真实验的准确性和可信度。

关键词: 网格仿真；Hadoop；数学模型

Massive Grid-Data Modeling Using Hadoop

HU Zhi-Gang, LIANG Xiao-Yang

(School of Information Science and Engineering, Central South University, Changsha 410083, China)

Abstract: Due to the needs of grid experiments and the shortcomings existed in the grid simulation tools, this paper presents an approach for massive grid data modeling using Hadoop technology. By using the modeling method in this paper, researchers can dig out the core data, which is required in the grid experiment, from the mass of data and then establish the mathematical model that the data needs. By using the experimental data generated by this model, the accuracy and credibility of grid simulation experiments will increase.

Keywords: grid simulation; Hadoop; mathematical model

1 引言

网格计算是分布式计算领域的一个重要分支。近些年来，在国内外研究人员的共同努力下，网格计算相关研究取得了长足的进展。由于网格计算本身所具有的高度复杂性(网格资源的动态性，不确定性，网格任务的异构性等)，网格计算理论的验证试验难度比较高[1]。此外，网格资源大都比较珍贵，网格研究人员在实际网格资源中进行理论验证的代价较高且周期较长，不利于研究的进行。因此，网格计算理论的验证试验成为研究网格计算的主要掣肘之一。

鉴于上述原因，众多研究机构都开发了自己的网格模拟工具对网格实验进行仿真。目前，几大主流网格模拟器有：GridNet、OptorSim、SimGrid、GridSim、ChicSim、EDGSim 等[2]。这些模拟器通过对网格资源，网格任务，网格用户等网格参与者的模拟以达到对真实网格环境的仿真。然而，大量实验结果显示：网格仿真实验结果与在实际网格环境中产生的实验结果相差较大，甚至与实际完全不符[3]。造成这一现象的主要原因是：实验人员采用的仿真实验数据(网格任务到达时间,作业长度等)大多是通过程序随机生成或者是某个网格节点一个时间段内的数据，而这些数据并不能很好的反映网格节点的实际情况。因此，如何选取符合实际情况的网格仿真实验数据成为网格研究领域亟待解决的问题。

2 Hadoop技术简介

Hadoop 是一个实现了 Google 的 MapReduce 计算模型的开源分布式并行编程框架，借助于 Hadoop，程序员可以轻松地编写分布式并行程序，将其运行于计算机集群上，完成海量数据的计算[4]。目前，企业界和研究机构都对 Hadoop 进行了深入的研究和应用。中科院高能物理研究所使用 Hadoop 技

① 基金项目:国家自然科学基金(60673165,60970038)
收稿时间:2010-02-26;收到修改稿时间:2010-03-26

术构建了集群进行海量数据处理；国内外各大公司如 FaceBook、Yahoo！、百度、淘宝等公司都使用 Hadoop 技术部署了自己的集群进行海量数据处理。Hadoop 技术已经成为海量数据处理研究方向的热门。

由于分布式存储对于分布式编程来说必不可少，Hadoop 框架中还包含了一个分布式文件系统 HDFS(Hadoop Distributed File System)，类似于 Google 的文件系统 GFS(Google File System)。HDFS 架构如图 1 所示：

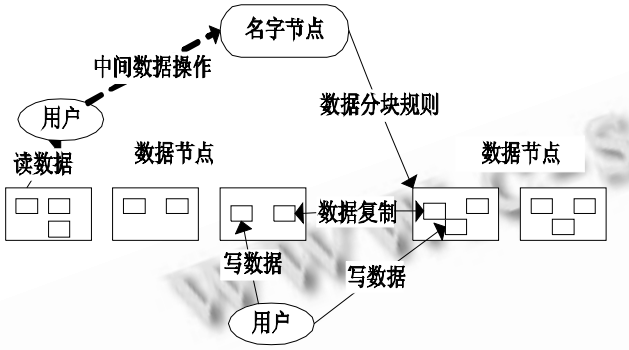

图 1　HDFS 架构图

HDFS 主要由 Client、Datanode 和 Namenode 组成。一个使用 Hadoop 技术架构的集群中，一般有一台主机作为 Namenode，若干台主机作为 Datanode。Client 代表使用 HDFS 的客户程序；Namenode 是 Hadoop 集群中的一台主机，负责保存数据节点的信息、计算任务的分发以及最终规约等任务；Datanode 负责数据存储与处理。为保证数据的安全性，HDFS 适度增加了冗余数据。具体的做法是在不同的 Datanode 中保存同一数据的三份拷贝，如图 2 所示：

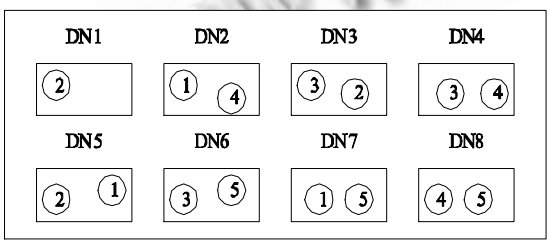

图 2　Datanode 示意图

图 2 所示的 Hadoop 集群中有 8 个 Datanode (DN1~DN8)，用于存储标号为 d1-d5 的 5 份数据。当数据节点 DN1 由于某种原因失效的时候，由于存储于该节点上的数据在其他节点上还存在备份，这样就保证了数据的安全性。

在数据处理过程中，Hadoop 采用 MapReduce 技术。如图 3 所示，MapReduce 分为 Map 和 Reduce 两个阶段。Map 阶段将要处理的任务依据某种规则划分为若干个阶段，Reduce 阶段将各个数据阶段的处理结果按照映射时候产生的键值进行规约。

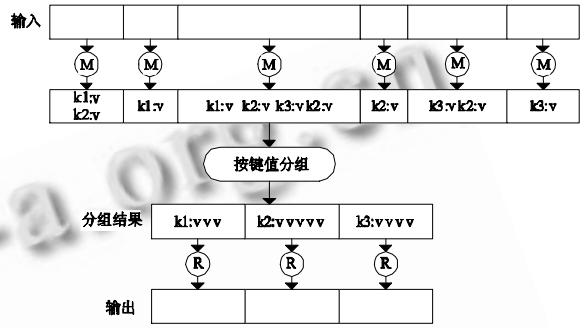

图 3　MapReduce 示意图

3　网格数据建模

在实际网格环境中，网格作业到达时间、本地负载等数据呈现一定的统计规律[5]。如果能够对这些数据服从的统计规律进行建模，然后根据建模产生的数学模型生成网格实验数据，即可提高网格仿真实验的准确性和可信度。然而，大型的网格节点每天都要处理大量的网格作业，生成的网格数据是海量级的。普通的 PC 机根本无法在短时间内处理这些数据，而使用大型计算机的代价较为昂贵，一般研究人员难以承受。

采用 Hadoop 技术，可以使用廉价的 PC 机搭建运算性能可观的集群系统在较短的时间内处理和分析海量数据。Hadoop 处理数据分为 Map 和 Reduce 两个阶段，因此使用 Hadoop 处理的数据必须易于分块且在逻辑上关联性较弱。网格数据正具备上述特性。

结合网格数据的特性和 Hadoop 的架构，本文提出的建模方法分为以下几个步骤：

(1) 选取统计数据，确定统计变量

在网格实验数据中，网格作业的到达时间、作业长度、完成时间以及网格资源预留率等因素对整个网格实验的最终结果影响较为明显。根据研究需要，选取其中之一作为统计变量，记为 e。统计变量的集合记为 E。

(2) 根据问题特征选择网格数据映射规则和归约规则

定义 1. 选取统计变量后，需要对统计变量代表

的数据按照一定规则进行处理，这种处理规则在本文中定义为映射规则，记为 M 。

根据选取的映射规则可将统计变量集合 E 划分为 n 个相互独立的子集 $E_1 \sim E_n$ 。

定义 2．实验数据按照映射规则进行处理后，将生成一系列中间结果，处理这些中间结果的方法在本文中定义为归约规则，记为 R 。

对于统计变量集 E 中的两个值 e_1 和 e_2 ，如果 $M(e_1) \in E_i, M(e_2) \in E_j$, 且 $i = j$, 则将 $M(e_1)$ ， $M(e_2)$ 归为一类，用归约规则 R 对二者进行归约处理。

(3) 实现 Mapper

按照映射规则实现 Hadoop 的 Mapper 接口，即覆写 map 方法。核心算法如图 4 所示：

```
map(Text inputData,OutputCollector<keyType,
      valueType> outputMap)
Step 1:Define a Pattern p;//定义输入数据解析规则
Step 2:E ← parse(inputData,p);//将解析值存入统计变量集 E
Step 3:for each e in E:
                for i=1 to n:
                      if M(e) ∈ Ei
                            collect<i,1> into outputMap;
                      else
                            collect<i,0> into outputMap;
                      endif
                endfor
        endfor
```

图 4　Mapper 实现算法

如图 4 所示，map 方法按照映射规则 M 生成用<key,value>形式存储的二元组结果集。如果映射值 $M(e)$ 属于某个统计变量子集，则将二元组的 key 值记为该子集的标号，value 值记为 1；反之，key 值不变，value 值记为 0。

(4) 实现 Reducer，对中间结果进行归约

按照归约规则实现 Hadoop 的 Reducer 接口，即覆写 reduce 方法。将第 3 步产生的中间结果进行归约。归约函数为：

$$R(< key_m, value_m >, < key_n, value_n >) = < key_m, value_m + value_n >, if\ key_m = key_n$$

即将结果集中 key 值相同的二元组的 value 值进行求和，生成新的二元组结果集。

(5) 定义 Hadoop 任务实例，启动运行

定义 Hadoop 任务后，将实现后的 Mapper 和 Reducer 类实例作为参数传递给该任务进行执行。

任务执行完毕，得到结果集 $S:\{< i, j >\}$ ， i 为统计变量子集的标号， j 为 i 号子集中元素个数。

(6) 建模

第 5 步得到的结果集是建模需要的核心数据。建模算法如图 5 所示：

```
Step 1:put all i in set S into a new set P;
Step 2:put all j in set S into another set Q;
Step 3:for all j in set Q:
                j ← j/sum(j);
          endfor
Step 4:f(x) ← curve(P,Q);
```

图 5　建模算法

在图 5 的第 4 步中，以集合 P 中数据作为自变量值，集合 Q 中数据作为因变量值，选择适当的随机分布曲线进行拟合，最终得出概率分布密度 $f_0(x)$ 。

(7) 假设检验

对第 6 步得到分布密度做如下假设检验：

$$H_0 : f(x) = f_0(x) \qquad H_1 : f(x) \neq f_0(x)$$

如果 H_0 成立，则接受该分布密度；反之，重复第 6 步，直到假设检验成立。

整个建模流程如图 6 所示：

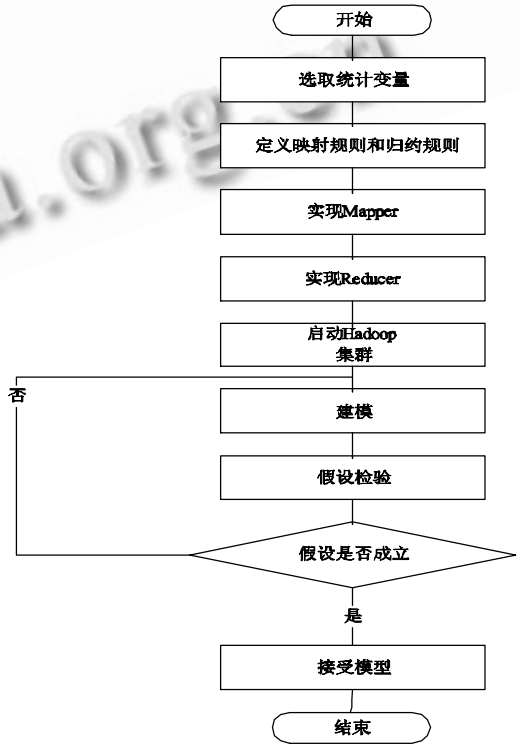

图 6　建模流程图

4　实验与分析

在本节中给出了一个使用 Hadoop 技术搭建的集群进行海量网格数据处理，并最终生成随机事件模型的实例来介绍使用 Hadoop 技术进行网格数据建模的示例。

本实验的软硬件环境如表 1 和表 2 所示：

表 1　实验硬件环境

主机编号	CPU 性能	内存大小	硬盘容量
0	Intel Pentium core2 E6300	2G	120G
1	AMD Sempron 2600+	1G	80G
2	AMD Sempron 2400+	1G	80G
3	Intel Pentium core2 E5400	2G	120G

其中，0 号主机的综合性能最佳。在采用 Hadoop 技术搭建的集群系统中，Namenode 的性能对整个集群的运算性能影响较大，因此选取该主机作为 Namenode，其他主机作为 Datanode。主机之间采用 1000M 交换机进行连接。

表 2　实验软件环境

操作系统	JDK	Hadoop	开发工具	数据分析软件
Ubuntu 9.10	JDK1.6	Hadoop0.19.0	Eclipse 3.5	MatlabR2007a

进行实验的四台主机都安装 Ubuntu9.10 操作系统。为了简化实验，采用统一的用户名和 SSH 密码。Hadoop 版本采用目前性能最为稳定的 0.19.0 版本，并安装在四台主机统一的系统目录下。搭建 Hadoop 集群，需要在 hadoop-site.xml 文件中加入如下内容，将本地文件系统配置成 HDFS，如图 7 所示：

```
<property>
  <name> fs.default.name </name>
  <value> hadoopserver:9000 </value>
</property>
<property>
  <name> mapred.job.tracker </name>
  <value> hadoopserver:9001 </value>
</property>
<property>
  <name> dfs.replication </name>
  <value> 2 </value>
</property>
```

图 7　HDFS 配置

本实例采用实际网格系统中产生的网格日志作为实验数据。网格节点在不同时刻可以提供的计算能力相差很大，所以网格任务的到达时间就成为影响整个网格节点性能的重要因素之一。使用本文提出的建模方法结合 Matlab 的曲线拟合工具，得出网格任务到达时间服从的概率分布密度为：

$y = f(x \mid a,b) = a * \exp(b * x)$，其中，

$a = 3.86 * 10^{-5}, b = 1.72 * 10^{-4}$，拟合结果如图8所示：

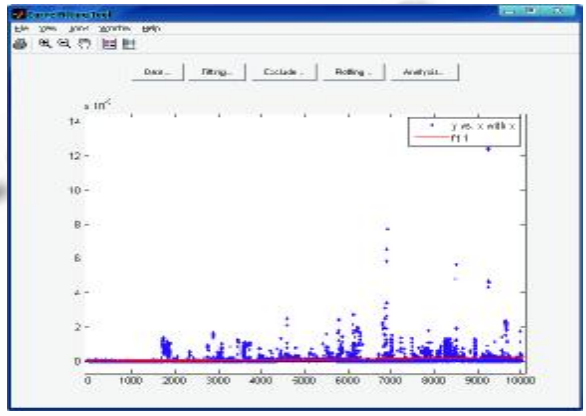

图 8　逼近结果图

对所得结果进行分布假设检验(本例采用卡方检验法，检验过程略)，经检验知，可以接受该模型作为任务到达时间的随机事件模型。

通常情况下，研究人员采用随机函数生成仿真实验数据。使用随机函数生成的网格任务到达时间大致上服从均匀分布。本例采用的实验数据中，不同时间段内到达的网格任务数相差很大，显然不服从均匀分布。可见，使用本文提出的建模方法生成的随机事件模型与实际情况更为接近。

5　结束语

本文介绍了一种结合 Hadoop 技术对海量网格数据建模的方法，并详细的阐述了该建模方法的各个步骤。最后，结合一个实例展示了如何使用本文提出的方法进行海量网格数据建模。

本文提出的建模过程分为利用 Hadoop 技术提取数据和利用 Matlab 分析数据两个过程，将 Hadoop 的海量数据分析功能和 Matlab 的数值计算和图像处理功能结合在一起为研究人员提供更为方便的仿真环境，将是下一步的主要工作。

(下转 17 页)

(上接 194 页)

参考文献

1 Sulistio A, Buyya A.A Grid Simulation Infrastructure Supporting Advance Reservation. Proc. of the 16th International Conference on Parallel and Distributed Computing and Systems (PDCS). Cambridge USA: ACTA Press, 2004.1 – 7.

2 Sulistio A, Cibej U, Venugopal S. A Toolkit for Modelling and Simulating Data Grids: An Extension to GridSim.Concurrency & Computation: Practice and Experience, 2008,20(13):1591 – 1609.

3 Jin H, Huang J. JFreeSim: A Grid Simulation Tool Based on MTMSMR Model. Advanced Parallel Processing Technologies-Lecture Notes in Computer Science, 2005,3756(13):332 – 341.

4 Apache Hadoop.Map/Reduce Tutorial.http://hadoop.Apache.org/common/docs/current/mapred_tutorial.html,2009.9 – 01.

5 Li H, Muskulus M, Wolters L. Modeling job arrivals in a data-intensive grid. Frachtenberg E, Schwiegelshohn U. Int'l. Workshop on Job Scheduling Strategies for Parallel Processing (JSSPP). Seattle, WA, USA: Springer, 2007.210 – 230.