

# 大数据技术在海量测震数据中的研究应用\*

郭凯<sup>1</sup>, 黄金刚<sup>2</sup>, 彭克银<sup>1</sup>, 庞丽娜<sup>1</sup>

(1. 中国地震台网中心, 北京 100045; 2. 山西省地震局, 山西 太原 030000)

**摘要:** 对目前主要的大数据技术进行探讨, 并对大数据计算方法在测震数据中应用进行分析, 设计了基于大数据技术的测震数据存储模型以及基于海量数据运行率管理的计算模型, 编写了分别基于单机多线程和大数据环境的程序, 并进行对比实验。结果表明: 采用基于大数据技术的分布式管理和计算架构, 可以较好地解决海量测震数据的管理困难、共享服务和科研计算方面的性能瓶颈。

**关键词:** 大数据; 测震数据; 分布式管理

**中图分类号:** P315.73

**文献标识码:** A

**文章编号:** 1000-0666(2017)02-0317-07

## 0 引言

大数据一般指大小超出常规的数据库获取、存储、管理和分析能力的数据集。大数据一般具有以下4个特征: 大量化(VOLUME)、快速化(VELOCITY)、多样化(VARIETY)和价值化(VALUE)。近年来, 大数据技术在计算性能、无限扩容等方面表现出巨大优势, 因此, 在医学、金融、科研领域已经得到了广泛的应用。

在“中国数字地震观测网络”项目完成后, 我国的地震监测能力得到了很大提高。相比“九五”期间, “十五”计划完成后中国国家地震台网中心的技术系统得到了全面提升, 无论在接入台站数量还是在数据处理效率上均有了较大的提升(侯建民等, 2009)。2008年至2016年底, 随着监测台网的不断完善和扩大, 存储的测震连续波形数据已经超过了100 TB, 每年还在以约12 TB的增量数据进行增加。如此海量的数据, 如何做好管理、服务、以及分析处理具有很大的挑战性。以前基于文件、MySQL数据库等方法进行管理的模式无论在数据管理、分析处理上都已经出现了很大的瓶颈, 时间和效率上已经无法满足现在地震科学数据管理和科研的需求。而随着大数据技术的高速发展, 其对海量数据的处理在IO并发、处理

速度上都体现了极大的优势, 山东省地震局(李永红等, 2015)、中国地震局第二监测中心(王丹宁等, 2016)、中国地震台网中心等都在该方面开展了相关研究工作, 本文主要探讨大数据技术在测震数据上的应用。

## 1 大数据技术发展

### 1.1 海量数据的存储技术

近年来, 随着数据量的高速增长, 对数据进行存储、备份以及分析时, 采用高性能集群和存储的成本非常高昂。Google公司于2003年提出了GFS的文件存储方法, 面对的数据规模是TB级或者GB级, 它采用价格低廉的存储和计算机进行稳定高效的海量数据管理和计算分析。2006年谷歌公司提出了面向结构化大数据的存储模型——Bigtable, 它是一个为管理大规模结构化数据而设计的分布式存储系统, 可以扩展到PB级数据和上千台服务器, 能提供灵活、高性能的存储解决方案。

Hadoop-HDFS(Hadoop Distributed File System)是文件分布式系统, 起源于Apache Nutch, 也是目前应用最广泛的大数据技术之一(Tom, 2014)。Hadoop在2006年2月成为一个独立的Lucene子项目, 是开放源码并行运算编程工具和分

\* 收稿日期: 2016-11-30.

**基金项目:** 地震行业科研专项“中国全球地震台网建设预研”(201508007)、国家自然科学基金“强噪声背景下极微弱信号的提取——低频可控震源核心科学问题之一”(D040901)和国家科技基础性工作专项“地球物理领域科技基础性工作数据资料集成与规范化整编(2013FY110900-2)”共同资助。

散式档案系统, 凭借其开源和易用的特性, 成为大数据处理的首选。HDFS 采用了主从 (Master/Slave) 架构, 一个集群有一个 Master 和多个 Slave, 前者称为名字节点 (NameNode), 后者称为数据节点 (DataNode), 放在 HDFS 上面的数据被分为一系列固定大小的数据块 (block) (许春玲, 张广泉, 2010)。Hbase 是一个高可靠性、高性能、面向列、可伸缩的分布式存储系统, 对于任何应用, Hbase 在逻辑上都把所有数据存储在一块表中 (许闯等, 2012)。

### 1.2 海量数据的快速计算

MapReduce (大规模数据集并行运算算法) 使得数据处理人员只需要执行简单的运算, 而将并

行化、容错、数据存储、负载均衡等数据处理细节放在一个标准库里 (Jeffrey, Sanjay, 2004)。图 1 为 MapReduce 的计算架构, 数据在 Map 阶段进行切割, 根据指定的 Key 组合为一个列表, 然后根据 Key 值进行分区, 在 Reduce 阶段进行计算合并。

Spark 是近年来发展较快的分布式并行数据处理框架。MapReduce 缺点在于 Map 和 Reduce 阶段产生的数据存储在硬盘上效率相对较低, 且只能处理离线数据, 面对需求实时处理的数据就无法满足需求。Spark 和 MapReduce 的区别之一在于 Spark 把中间产生的数据放在了内存中, 并采用 RDD (Resilient Distributed Datasets, 弹性分布式数据集) 来提高计算效率。

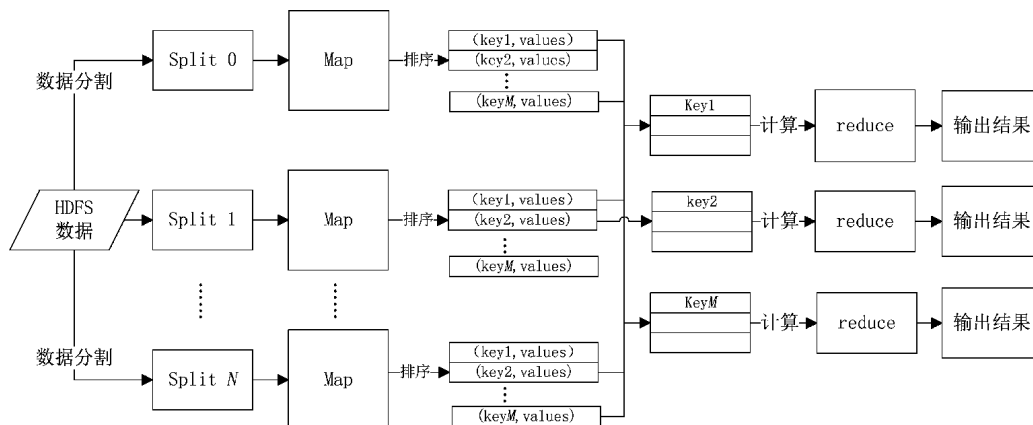


图 1 MapReduce 计算架构

Fig. 1 Computing architecture of MapReduce

## 2 大数据技术在测震数据中的应用

IRIS DMC 是全球地震台站最大的数据汇集中心和处理中心, 世界各个地震研究中心和研究机构通过 Wilber、SeedLink、ArcLink、WebService 等标准传输方式从 IRIS DMC 获取数据。图 2 为 IRIS DMC 在 2006—2016 年期间数据汇集和数据服务情况, 从图中可以明显看到汇集的数据量不仅从 2006 年的几十 TB 增长到现在将近 400 TB, 对外服务提供的数据量更是呈现爆发式增长, 2016 年的数据服务量比 2006 年增长了将近 50 倍, 达到了 900 TB。显然, 如果采用普通的数据管理和服务方式, 尤其是在破坏性大地震发生后面对全球用户的大量数据请求响应和处理下载, 是无法满足规模日益增长的数据服务需求的。

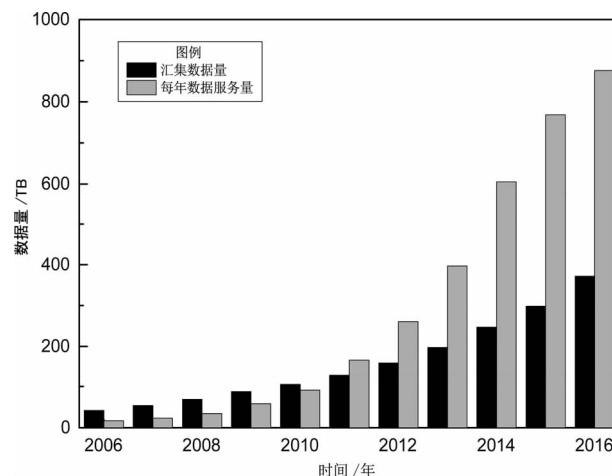


图 2 IRIS DMC 数据汇集和数据服务量 (2006—2016)

Fig. 2 Amount of data collection and service of IRIS DMC (2006—2016)

中国地震台网中心从最初接入的 47 个国家标准数字地震观测台站（赵永等，2002），到已经实现了 1 024 个国家和地区地震台站的测震波形数据实时传输和汇集（郭凯等，2016），监测能力大幅提升的同时，需要实时处理和存储的数据量也达到了之前 20 多倍。以前将数据备份到光盘、磁带库的方式在急速增长的数据量面前暴露出数据易损坏、恢复难、数据导入导出速度慢等诸多问题。从数据服务的角度，由于国家测震台网西部分布相对稀疏，东部分布密集。如果  $M5.0 \sim 6.0$  地震发生在西部，震后时间波形数据由于台站密度相对稀疏，需要处理的台站数据相对较少。但如果发生在华北区域，按震中 800 km 选取地震台站 30 min 数据，最少需要处理 300 个台站产生约 0.5 GB 的数据，面对较多用户的数据处理和下载请求时，对系统 IO 和并发处理能力提出了非常高的要求，普通服务器很难满足。

2.1 基于 Hadoop 的测震数据存储模型架构设计

中国地震台网中心汇集的测震波形数据采用国际标准的 Miniseed 格式，以一个台站一个分项 512 字节、每个数据包 1 天 24 个小时数据做为一个文件的形式存储，目前所有的数据汇集在 NAS 存储上，受限于网络带宽以及 NAS 机头数量，传输速度很难超过 100 M/s。在面对 TB 级规模的测震波形数据，单纯采用文件方式存储在数据汇集的速度、稳定性和安全性方面已经无法满足要求。表 1 为目前传统主要使用的存储和云存储特性比较。

表 1 传统存储和云存储的特性比较

Tab. 1 Comparison of the characteristics of traditional storage and cloud storage

特性	NAS 存储	SAN 存储	云存储/分布式
扩展性	传统 NAS 机头不可扩展	扩展性差	在线扩展
可靠性	机头单点故障	单点故障	无单点故障，高度可靠
管理性	难以管理	难以管理	简捷易用
并发性	性能受限于机头	性能优于 NAS	随系统规模增加呈线性增长

目前国内已开展的地震大数据存储方式主要分为 2 种：①将数据按照原始格式导入 Hbase 中（王丹宁等，2016）；②对原始测震波形数据进

行解压缩，整理成 ASCII 码放入 Hbase 中。由于 Miniseed 格式本身采用了 Stein2 的压缩算法，直接将其放入 Hbase，对于提取数据时的计算效率会有一定的影响，而将数据解压成 ASCII 码的形式，则增长的数据量达到了将近 4 倍，会造成很大的存储消耗。从测震数据的使用需求来看，主要分为 2 种：①实时性计算，主要用于地震速报、地震预警和烈度速报等；②数据分析，主要对历史数据进行计算分析，如噪声成像、区域速度结构等科学研究领域。由于 HDFS 是为了处理大型数据集分析任务的，是为达到高的数据吞吐量而设计的，这就可能要求以高延迟作为代价。

基于测震数据的实际业务需求和汇集情况，并且考虑到数据存储的成本，本文提出了如图 3 所示的基于 Hadoop 的测震数据存储架构。将实时数据和近期 1 年的测震波形数据放入 Hbase 中存储，这样可以满足对实时性计算要求较高的地震速报和地震预警要求；将历史数据放入 HDFS 中，并采用 1:3 的比例进行数据备份，提高了数据的安全性，并可以开展基于 MapReduce 和 Spark 的高效计算。

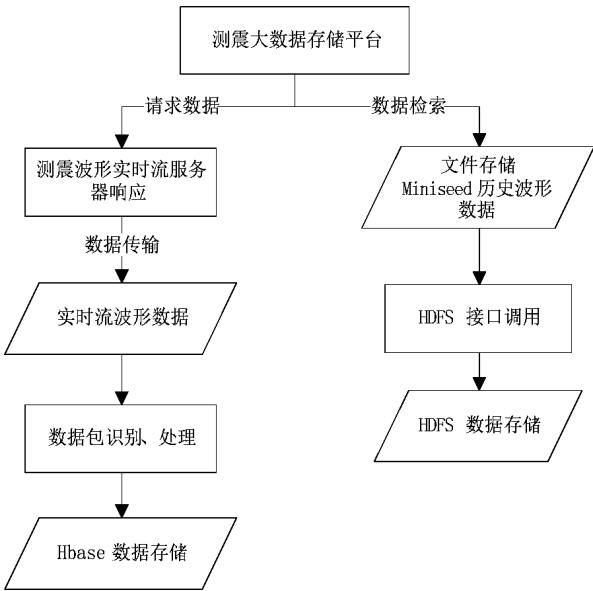


图 3 基于 Hadoop 的测震数据存储模型  
Fig. 3 Storage model of seismic data based on Hadoop

当然，测震数据在 Hbase 中的存储要结合实际的业务需求进行设计。目前国家台网中心采用基于 JOPENS 的 SSS 流服务器来负责接收和分发近实时波形数据，接收数据为 512 字节的纯数据 SEED

卷 (Miniseed 数据), 包含有固定头段部分 (48 字节) 和数据部分, 主要记录台站名、通道名、记录起始时间、样本数目、测震数据等, 以一个台站一个分项 (周辉等, 2011)。刘坚等 (2015) 对测震数据存入 Hbase 的数据结构进行相关设计和研究, 设计了 Row Key 为 < Netid > < Stationid > < Pointid > < Intrid > < Itemid > < Samplerate > < Timestamp > < Prototype >, 主要基于 512 字节数据头段所包含的台站信息来进行数据的快速检索 (刘坚等, 2015)。根据台网中心的测震波形文件存储格式, 并考虑实际业务需求和数据检索效率, 本文设计了如表 1 所示的基于 Hbase 的测震波形数据存储格式。考虑性能和维护性, 采用每月数据存储在设计思想, 每张表有 2 列, 一列是 ROWKEY, 另一列存储每条实时流, 以北京 BST 台站 2016 年 4 月 9 日 16 点 10 分 32 秒接收到的数据为例, 设计的表结构如表 2 所示。

表 2 基于 Hbase 的测震波形文件存储格式

Tab. 2 Seismic data format based on Hbase

ROWKEY	VALUE
月份_时间戳_台网代码_台站代码_测项_分项_采样率	数据包
04_2016.110.08.10.32_BJ_BST_00_BHZ_100	512 字节数据
04_2016.110.08.10.32_BJ_BST_00_BHE_100	512 字节数据
04_2016.110.08.10.32_BJ_BST_00_BHN_100	512 字节数据

ROWKEY 设计的月份\_时间戳\_台网代码\_台站代码\_测项\_分项\_采样率, 可以从实时流中的数据头段解析而来, VALUE 为整个实时流数据。由于很多测震数据的研究往往以台网为单位进行数据处理, 为了进一步提高数据的检索效率, 可以基于台网编码建立二级索引来进一步提高数据的检索效率, 但在提高效率的同时要付出消耗更多空间的代价。

## 2.2 数据管理和计算模型设计

测震波形数据质量控制是数据共享和服务非常关键的一步, 而运行率是评估台站数据质量在一定时间范围内的一个重要指标。如果以 5 年全国测震台站波形数据 1 024 个台站约 65 TB 数据, 每个台站 1 个分项 1 小时运行率作为基本单位进行统计并存储该小时段的运行率, 则产生的数据量超过了 2 亿多条, 而传统的数据库如 Mysql 在规模达

到几百万条数据的时, 检索的速度就已经达到了瓶颈, 同时, 计算如此海量的数据, 即使采用高性能计算机进行数据运行率检索, 也会受限于 IO 瓶颈和 CPU 的数量, 计算的时间会非常漫长。

基于台网中心目前的数据存储情况以及本文设计的测震数据存储模型, 设计了一套基于 Hadoop 的分布式历史数据运行率检索计算模型, 同时兼容考虑了对实时流数据运行率的实时检索, 如图 4 所示。模型基于表 2 的数据存储结构将历史数据首先导入 HDFS 中, 采用 Spark 计算模式开展多节点的数据运行率计算, 以小时为周期将计算结果实时放入 Hbase 中, 实时流数据的计算流程同历史数据一样, 结果放入 MySQL, 并按一定周期导入 Hbase 中。

图 4 中, 自底向上分为 3 个层面。第 1 层是数据源层, 主要包括地震波形数据文件以及测震实时流, 最新的地震波形数据文件存储在 NAS 存储服务器上, 可通过脚本将其挂载到本地进行读写, 历史地震波形数据文件存储在 HDFS 分布式文件系统中, 用 HDFS 接口进行存储和访问; 第 2 层是处理层, 主要完成系统所需要数据计算功能, 主要提供基于历史地震波形数据文件的连续率计算以及实时地震波形数据文件的连续率计算; 第 3 层是云平台中系统的数据持久层, 主要提供处理层中各类计算结果的存储, 包括测震连续波形数据、索引数据以及台站通道信息数据等。其中, 连续率数据由于数据量巨大, 将存储在分布式数据库 Hbase 中, 通过 Hbase 接口进行存储和访问处理; 而其它类似基础信息类的数据, 即数据量不大但处理响应性能要求较高的数据, 将存储在关系数据库系统 Mysql 中, 用 JDBC/SQL 进行存储和访问处理。

## 2.3 基于大数据架构的测震数据计算测试

本次测试编写了 2 个测震数据运行率计算程序, 分别为基于 Hadoop 的大数据计算版本和基于多线程的单机版本。这里主要介绍大数据计算版本的程序设计和执行步骤: 首先数据处理模块的主线程获取 HDFS 上的数据目录集合, 并将目录集合以任务集的方式提交至计算节点进行计算, 计算节点每次取出一天的数据进行计算, work 从 HDFS 上取出某天目录下的所有 Miniseed 文件依次进行解析, 每个 Miniseed 文件按 512 个字节为单位进行读取, 读取时需要对重复数据进行选优计算,

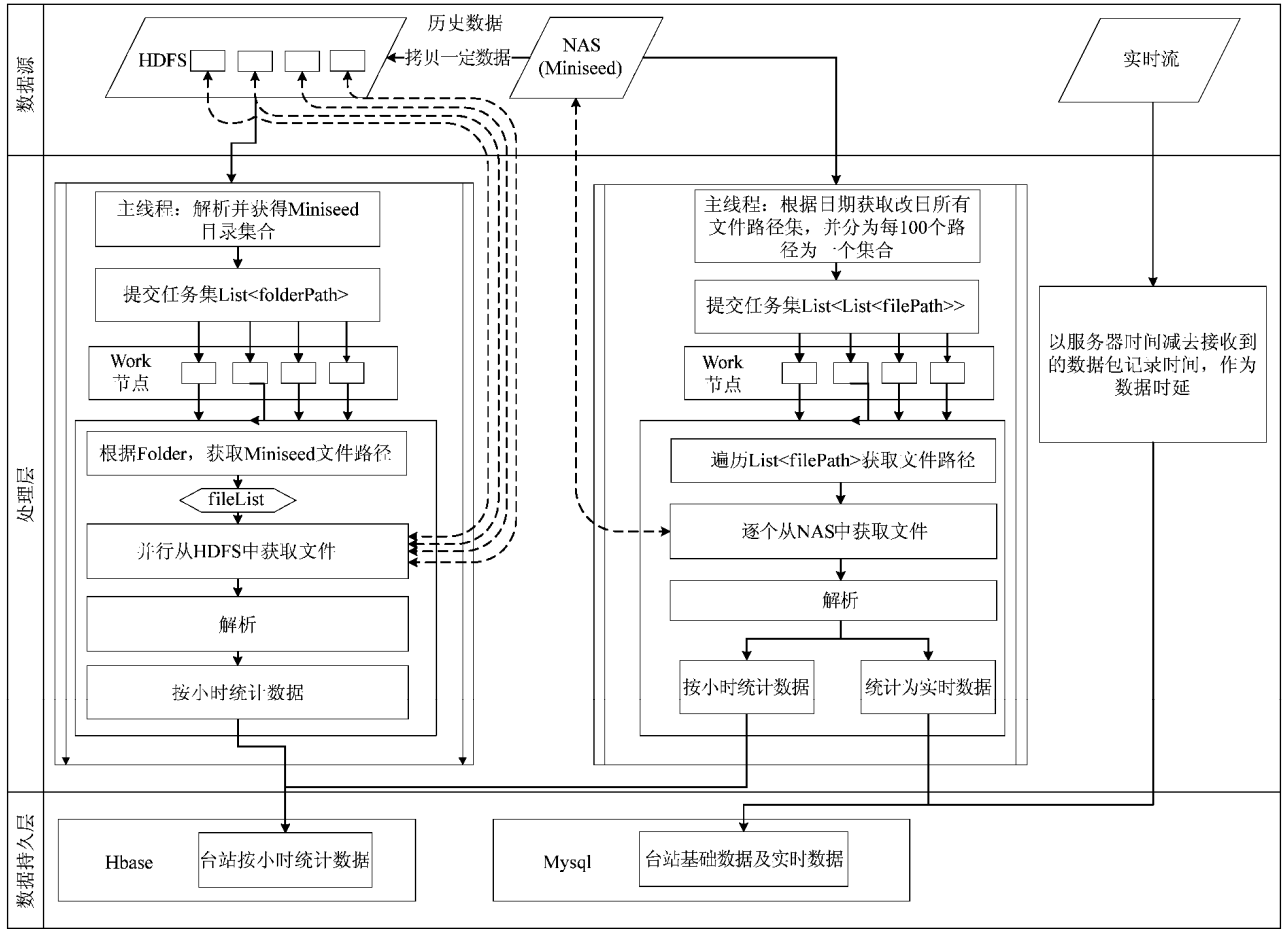


图 4 基于 Hadoop 的测震数据运行率计算模型

Fig. 4 The calculation model of the operation rate of the seismic data based on Hadoop

并按每个小时为 key， $n$  个 Miniseed 数据为集合对 value 生成一个该日的 HashMap；接下来按每小时为单位遍历该日的 HashMap，读取 Miniseed 数据的台站信息并计算该小时的连续率；最后每个 Miniseed 文件计算完毕后，将解析出台网代码、台站代码、连续率计算结果一起推送到数据存储层进行存储。

测试采用 6 台服务器组成的集群进行测试，集群中每台服务器的软硬件情况如表 3、4 所示。

表 3 服务器硬件配置		
Tab. 3 Hardware configurations of the servers		
类别	详细配置	数量
主板	X9DRI - F	1
处理器	Intel Xeon E5 - 2650 - V2 + EG - 2UTDL2011	2
内存	镁光 16G DDR3 - RECC 1600	8
数据盘	WD 4TB SATA 3.5	12

表 4 计算集群环境配置				
Tab. 4 Environment configurations of computing cluster				
IP 地址	节点类型	Zookeeper	Kafka	Spark
10. 5. 190. 52	Master	无	无	
10. 5. 190. 53	Second, Slave	无	无	
10. 5. 190. 54	Slave	无	无	Master
10. 5. 190. 55	Slave	无	有	Worker
10. 5. 190. 56	Slave	有	有	Worker
10. 5. 190. 57	Slave	有	无	Worker

ZooKeeper 是一个开放源码的分布式应用程序协调服务，是 Hadoop 和 Hbase 的重要组件，提供的功能包括配置维护、域名服务、分布式同步、组服务等；Hbase 是一个分布式的、面向列的非结构化开源数据库，在纵向上可以提供无限扩展能力；Kafka 是一种高吞吐量的分布式发布订阅消息系统，它通过 Hadoop 的并行加载机制来统一各类消息处理，也是为了通过集群来提供实时数据的

缓存和消费; Spark 是一种与 Hadoop 相似的基于内存的开源集群计算环境, 由于 Spark 启用了内存分布数据集, 除了能够提供交互式查询外, 它还可以优化迭代工作负载。

测试数据采用 2015 年 12 月的全国测震波形数据, 数据量约 1 TB。多线程版本测试将 1 个月的 Miniseed 文件保存在 10.5.190.56 服务器的硬盘上, 目录按每天 1 个文件夹进行存放, 通过多线程技术分别将 1 个月的文件进行解析, 测试结果如表 5 所示。

表 5 多线程程序计算结果

Tab. 5 Calculation results of multithreaded program

测试用例	线程数	文件个数	测试所用时长
1 月文件解析	3	87 800	354 min

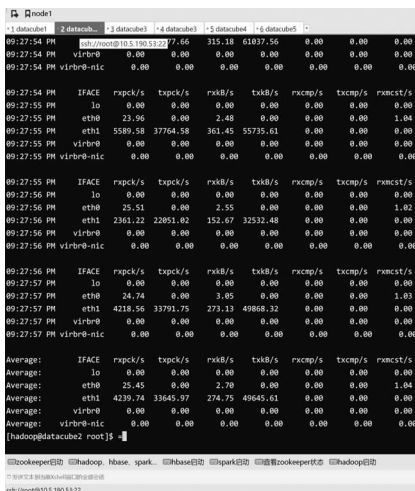
大数据版本测试将 1 个月的 Miniseed 文件保存在 10.5.190.56 服务器的硬盘上, 目录按每天 1 个文件夹进行存放, 通过执行 Hadoop 的数据导入命令, 将这些数据上传至 HDFS 文件系统中, 执行 Spark 程序进行测试, 测试结果如表 6 所示 (Spark 配置使用 15 个 cpu 内核, 以 1 个任务进行提交)。

表 6 大数据程序计算结果

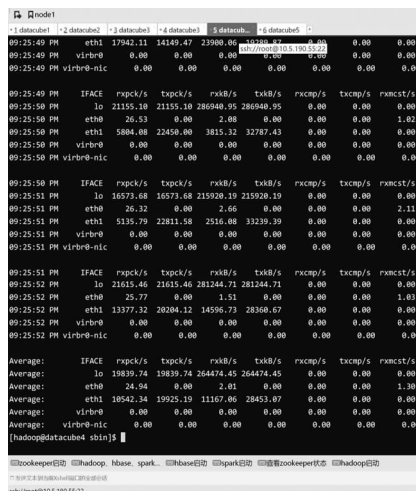
Tab. 6 Calculation results of big data program

测试用例	cpu 核心数	文件个数	测试所用时长
1 月文件解析	15	87 800	53 min

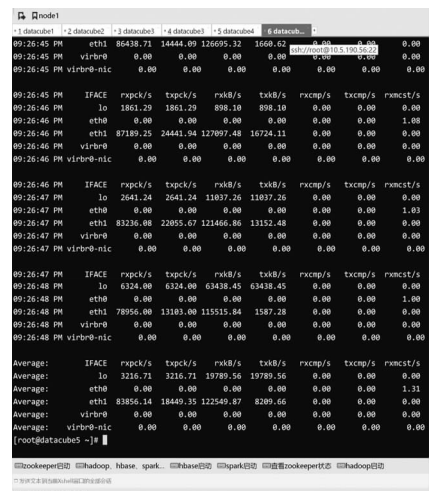
通过上面测试, 发现大数据版本的处理时间仅为多线程版本的 1/6, 多线程版本的数据放在本地硬盘, 受限于文件读取效率, 而从 HDFS 文件系统读取, 不限制于硬盘读取速率, 且可以成倍的提高读取解析效率; 另外, 如图 5 所示, 其中计算节点 10.5.190.53 (图 5a) 上只有 HDFS, 所以上传速率是 30~40 M/s, 下载速度为 0, 表示该节点在为 spark 运算提供文件数据; 计算节点 10.5.190.55 (图 5b) 上不仅有 HDFS 节点, 还有 Spark 计算节点, 上传和下载速度约 264 M/s, 表明该节点不仅有 Spark 的 work 节点在接受并解析数据, 同时, 该节点在向外发送数据; 计算节点 10.5.190.56 (图 5c) 上也是不仅有 HDFS 节点, 同时有 Spark 计算节点, 但是上传速率很低, 说明该解析的数据并没有存放在该节点上或存放量偏少。但是该节点文件接受速率为 120 M/s, 说明该节点在解析数据。由于本次测试的 Hadoop 集群配置的是千兆网卡, 因此, 120 M/s 左右的速率为正常速率, 220 M/s 的上传下载速度表明该节点不仅接受其他 HDFS 节点传输过来的数据, 同时在处理本机上的 HDFS 节点数据。综上可知, 通过 Spark 计算, 还有一定提升空间, 由于网络带宽达到上限, 通过增加服务器的方式, 对网络进行负载均衡, 可以进一步提升 Spark 解析任务的文件读取效率, 进一步提升解析速度。



(a) 计算节点 53 运行界面



(b) 计算节点 55 运行界面



(c) 计算节点 56 运行界面

图 5 基于 Hadoop 的集群计算节点运行界面

Fig. 5 Running interface of computing cluster base on Hadoop

### 3 结论

本文从测震数据存储和共享科学计算的角度出发,就大数据技术在海量测震数据的存储和应用方面进行了相关研究,基于海量测震数据的实际业务需求和汇集情况,设计了一套分别将测震数据放入 HDFS 和 Hbase 的数据模型,并通过多副本的设定来保证数据的安全性;从测震数据质量控制的需求出发,提出了一个基于 Hadoop Spark 的海量数据运行率计算模型,采用测震数据就传统的多线程计算和基于 Hadoop 环境的集群计算做了对比实验,实验结果体现了大数据在海量数据处理上进行分布式存储和计算的强大的性能优势,该性能可以随着参与计算节点的增加来进行提升。因此,对于开展地震波形互相关、层析成像、区域速度结构等研究,需要对海量测震波形数据分析处理的地震科研人员来说,可以大大提高效率,使处理速度达到之前的几倍、几十倍甚至更多,体现了大数据技术在地震行业的利用价值。

本文在撰写过程中得到孟令媛副研究员、南

京云创大数据公司马鸣、汪洲权的帮助,在此向他们表示衷心感谢。

### 参考文献:

- 郭凯,温瑞智,杨大克,等. 2016. 地震预警系统的效能评估和社会效益分析[J]. 地震学报,38(1):146-154.
- 侯建民,黄志斌,余书明,等. 2009. 中国国家地震台网中心技术系统[J]. 地震学报,31(6):684-690.
- 李永红,周娜,赵国峰,等. 2015. 云计算环境下地震数据管理与服务应用研究[J]. 震灾防御技术,10(增刊1):811-817.
- 刘坚,李盛乐,戴苗,等. 2015. 基于 Hbase 的地震大数据存储研究[J]. 大地测量与地球动力学,35(5):890-893.
- 王丹宁,柴旭超,王文青. 2016. Hadoop 平台下的地震波形数据存储与应用规划[J]. 软件工程,19(1):48-49.
- 许闯,刘鹏,刘志忠,等. 2012. 一种基于 HBase 的本地复用新方法[J]. 计算机技术与发展,22(6):57-60.
- 许春玲,张广泉. 2010. 分布式文件系统 Hadoop HDFS 与传统文件系统 Linux FS 的比较与分析[J]. 苏州大学学报(工科版),34(1):5-9.
- 赵永,薛峰,刘阳,等. 2002. 国家数字地震台网中心技术系统与服务[J]. 地震地磁观测与研究,23(1):16-23.
- 周辉,申学林,王文青,等. 2011. 通用测震数据获取软件包的设计与实现[J]. 地震研究,34(1):102-107.
- JEFFREY D, SANJAY C. 2004. MapReduce: Simplified Data Processing on Large Clusters[C]. Conference on Symposium on Operating Systems Design & Implementation, 137-150.
- TOM W. 2014. Hadoop 权威指南[M]. 北京:清华大学出版社.

## Research and Application of Big Data Technology in Mass Seismic Data

GUO Kai<sup>1</sup>, HUANG Jingang<sup>2</sup>, PENG Keyin<sup>1</sup>, PANG Lina<sup>1</sup>

(1. China Earthquake Networks Center, Beijing 100045, China)

(2. Earthquake Administration of Shanxi Province, Taiyuan 030000, Shanxi, China)

### Abstract

In order to solve the bottlenecks of management, service and the calculation with the growing mass seismic data, we research the main large data technology and analyze the application of large data calculation method in seismic data, design the storage model of seismic data based on large data technology and the calculation model based on mass data operation rate management, at last we make a comparison experiment between single machine multi-thread and large data environment. The results show that the big data technology of distributed management and computing architecture can solve the bottlenecks problem of the management, data service and scientific research of mass seismic data.

**Keywords:** big data; seismic data; distributed management