

单维锋、滕云田、刘海军等,2019,大数据环境下地震观测数据存储方案研究,中国地震,35(3),558~564.

大数据环境下地震观测数据存储方案研究

单维锋^{1,2)} 滕云田¹⁾ 刘海军²⁾ 杨冠泽²⁾

1) 中国地震局地球物理研究所,北京 100081

2) 防灾科技学院,河北三河 065201

摘要 大数据技术为处理海量地震观测数据提供了一种新的数据存储与计算模式。为了解决现有基于关系数据库存储方案的读写速度低、用户并发度低和可扩展性差等问题,以地震前兆观测数据为例,在详细分析业务需求的基础上,提出了基于 HBase 和 OpenTSDB 的地震大数据存储方案,搭建了大数据测试平台,完成了不同存储方案下查询、插入性能实验和并发性实验。实验结果表明,与关系数据库存储方案相比,基于 HBase 和 OpenTSDB 的存储方案具有很好的可扩展性和并发性,经过优化后的 HBase 存储方案具有更高的读取和存储性能。

关键词: 大数据 地震 Hadoop HBase OpenTSDB 并发

[文章编号] 1001-4683(2019)03-0558-07 [中图分类号] P315 [文献标识码] A

0 引言

近年来,随着监测仪器数量的增多和采样率的提高,地震观测数据量激增。现有基于关系数据库的存储方案存在读写速度慢、用户并发度低、可扩展性差等问题,难以满足科研人员日益增长的快速数据处理的需求(刘坚等,2015)。大数据技术为处理海量地震观测数据提供了一种全新的数据存储与计算模式。高效、合理的地震观测数据存储是后续开展机器学习、数据挖掘和地震分析预测等应用的前提和基础。本文以地震前兆观测时间序列数据为例,探讨了地震大数据存储方案(郭凯等,2017;李永红等,2015,)。

刘坚等(2015)分别针对地震结构化数据与非结构化数据提出了一种基于 HBase 的地震大数据存储方案,并与基于 MySQL 的存储方案在数据读写性能方面进行了比较研究。王丹宁等(2016)针对测震波形数据,分析了实时数据、历史数据的不同业务需求,初步提出了一种基于 HBase 的分布式存储方案。虽然前人在地震大数据存储方面开展了一些研究,但大多是基于 HBase 给出的较为笼统的设计方案,对地震前兆观测数据为时间序列数据的特点和常用的业务场景考虑不够全面,且仅仅测试了单用户模式下系统的读写性能。由于大数据环境下,数据趋于集中,并发需求大,因此对不同存储方案的并发读写性能测试尤为重要。

本文以地震前兆观测数据为例,在深入分析地震数据处理业务需求的基础上,结合地震

[收稿日期] 2018-11-06; [修定日期] 2019-07-01

[项目类别] 地震科技星火计划项目(XH16059)、河北省重点研发计划(16210126)、中央高校基本科研业务费项目(ZY20160106)、国家重点研发计划(2018YFC1503806)共同资助

[作者简介] 单维锋,男,1977年生,教授,博士,主要从事地震大数据技术研究。E-mail:shanweifeng@cidp.edu.cn

时间序列数据的特点和业务需求,首次将 OpenTSDB 通用时间序列数据库应用于地震前兆观测数据,然后提出了基于 HBase 的优化设计方案,并重点测试和分析了基于 MySQL、OpenTSDB 和 HBase 的地震前兆观测数据存储方案的读写性能和并发性能。

1 地震前兆观测数据的数据特点

地震前兆观测数据主要包括流体、形变、地磁、地电等学科数据,是典型的时间序列数据,现存储在关系数据库中。目前,正常运行的地震前兆观测仪器近 3000 套,测项分量近 8000 个,年产出数据量约 500G 左右(周克昌等,2013)。不同的仪器具有不同的采样率,包括秒采、分采、时采、日采等仪器。在实际应用中,由于经常需要对某个具体仪器的观测数据进行读取、分析和处理操作,因此将相近测项的数据合并存储在一张表中,同时将某台仪器产生的观测数据以天为单位压缩为 1 个字符串,存储在 1 条记录中,大大减少了表中记录的总条数,提高了数据查询效率,但在一定程度上增加了程序逻辑的复杂度。

2 地震业务处理特点分析

地震前兆观测数据是典型的时间序列数据,这些观测数据是专家进行地震趋势判断、数据分析与挖掘、地震预报的基础。在实际业务处理中,绝大部分数据读取业务是按时间段连续读取某台观测仪器的观测数据,几乎没有随机读取操作即通过组合台站、测点、测项、采样率和起止时间等条件查询数据。以下是几个典型的地震前兆原始数据处理场景:

(1) 读取某台仪器某个测项在某一段时间内的连续数据。

(2) 读取多台仪器多个测项在某一段时间内的连续数据。为了对比分析不同观测数据的变化趋势(如判断前兆异常),需要对一段时间内多台仪器的多个测项连续观测数据进行读取、处理,以对比发现其中可能存在的规律。

(3) 聚合函数应用。由于数据可视化、分析等业务处理的需要,通常需要对某台仪器的某个测项的某段数据内的数据进行聚合操作,如取平均值、最大值、最小值等。

3 大数据环境下地震前兆观测数据存储方案

地震前兆观测数据为时间序列数据,属于结构化数据范畴。由于现有的关系数据库在大数据规模下的读取性能和并发性能等方面难以满足科研人员需要,在设计大数据存储方案时,必须考虑地震前兆观测数据时间序列数据的固有特点以及按时间段读取、聚合的业务特点。目前,地震前兆观测仪器以分采和时采设备为主,秒采设备和日采设备较少。由于地震前兆观测数据存储平台需要为用户提供数据查询、分析、可视化服务以及为各类异常事件检测、地震分析预报等算法提供原始数据,需要高效读取数据,因此需要高效的并发读取性能。

在大数据环境下存储地震前兆观测数据主要有 3 种方案,一是直接基于 HDFS 文件系统进行存储,可按照台站、测项、仪器等建立多级目录,然后将观测数据按时间片段(如 1 天)进行存储。该方案简单、可靠、有效,但在读取长时间段或多个仪器数据时,需要读取多个数据文件,此时需要编写业务逻辑进行数据合并,比较繁琐。二是以 OpenTSDB 为代表的时序数据库。OpenTSDB 作为分布式、可伸缩的通用时间序列数据库,能够提供毫秒级精度

的时间序列数据存储,并提供了多种简便的访问接口,无需过多设计,可以拿来即用。三是以 HBase 为代表的 NoSQL 数据库。HBase 是高可靠性、高性能、可伸缩的列式数据库,可以支持超大规模数据存储,但需要用户根据数据特点和业务需求设计数据存储结构,难度较 OpenTSDB 数据库大很多。本文以地震前兆观测大数据为例,基于 OpenTSDB 和 HBase 数据库分别设计了地震大数据存储方案,并对其读写性能进行了对比分析研究。

3.1 基于 OpenTSDB 的数据存储方案

OpenTSDB 面向时间序列数据,提供了一套通用的时间序列数据存储和查询方案。在 OpenTSDB 中定义时间序列数据仅需要包含以下属性:指标名称 (metric name)、时间戳 (timestamp,单位为 ms 或者 s)、值 (value,64 位整数或者单精度浮点数)、1 组标签 (tags,用于描述数据属性,每个标签由 tagKey 和 tagValue 组成)。

OpenTSDB 底层依赖 HBase 数据库,但在存储上做了大量的优化,如 OpenTSDB 为每个 metric、tagKey 和 tagValue 都分配了 1 个 UID(固定为 3 个字节),大大缩短了 row key 长度,节省了存储空间。将同属于 1 个时间周期内(默认 1h)的具有相同 TSUID(相同的 metric 以及相同的 tags)的数据合并为 1 行存储,大大减少了时间序列的行数,提高了查询效率和存储效率。本次测试的存储方案如表 1 所示。

表 1 基于 OpenTSDB 的存储设计方案

指标名称	时间戳	值	标签
测项	观测日期(精确到日)	具体的观测值	台站、测点、采样率

3.2 基于 HBase 的数据存储方案

HBase 是一个高性能的列式数据库,它可以处理超过 10 亿行数据和数百万列数据组成的数据表,其关键在于设计合理的 row key,以方便数据查询。与刘坚等(2015)提出的 row key 设计方案不同,我们借鉴了 OpenTSDB 设计思路,将观测仪器 1 天内的观测数据合并为 1 条记录,以采样率为列簇,以该采样率 1 天内包含的数据数目为列数,如 Seconds 列簇有 86400 列,Minutes 列簇有 1440 列,Hours 列簇有 24 列,Days 列簇只有 1 列。row key 为“台站+测点+测项+采样率+时间(日)”,其中“时间”精确到日,如表 2 所示。

表 2 基于 HBase 的存储设计方案

row key	列簇:Seconds	列簇:Minutes	列簇:Hours	列簇:Days
台站+测点+测项+采样率+时间(日)	列:0, ..., 86399	列:0, ..., 1439	列:0, ..., 23	列:0

此方案的优点在于通过合并 1 天内的观测数据为 1 条记录,一方面可以大大减少表中记录的总数目,提高查询效率;另一方面,通过将 1 天内的数据存储在不同的列中,可直接读取某列来获取其值,不需要字符串拆分操作;最后,通过离线计算秒采、分采和时采数据的分均值、时均值和日均值数据等,并写入到对应的列中,大大提高部分聚合查询请求的性能。

4 实验对比

4.1 实验环境

为对比大数据环境下地震前兆观测数据不同存储方案的存储、查询性能指标,基于 OpenStack 云平台搭建了由 10 台服务器组成的 Hadoop 高可靠(HA)集群。每个节点包含 1 颗 2 核 CPU、8G 内存和 500G 硬盘。Hadoop 集群的每个节点均安装了 CentOS 7 和 JDK 1.8,并在相应的节点上安装了 Hadoop 2.7.6、zookeeper 3.4.10、HBase 1.3.2、和 Opentsdb 2.3.0 软件。每个节点的角色分配情况如表 3 所示,表中“√”表示本节点安装了该项配置,M 表示 HMaster 节点,S2 为 HMaster 备份节点,R 为 RegionServer 节点。8 台 RegionServer 节点上均部署了 OpenTSDB 服务。

表 3 HBase 与 OpenTSDB 的实验配置

节点	Namenode	Datanode	Journalnode	Zookeeper	ZKFC	RM	Hbase	OpenTSDB
S1	√				√	√	H	
S2	√			√	√	√	H	
S3		√	√				R	√
S4		√		√			R	√
S5		√	√				R	√
S6		√		√			R	√
S7		√	√				R	√
S8		√		√			R	√
S9		√					R	√
S10		√		√			R	√

此外,为了测试关系数据库的系统性能,选用 1 台硬件配置较高的服务器进行单机测试,该主机包含 2 个 Intel(R) Xeon(R) E5-2620 V4 CPU(2.1 GHz,8 核),64G 内存和 5 块 2T 硬盘(RAID 5)。安装了 CentOS 7、JDK 1.8 和 MySQL 5.0 数据库,并创建了现有数据存储方案。

所有测试指令由 1 台单独的 PC 计算机执行,并发测试采用多线程技术进行模拟实现。

测试程序应用 Java 开放语言,程序通过 JDBC 连接 MySQL 数据库实现读写操作,通过 Native Java API 访问 HBase 集群,通过 HTTP API 接口连接到 OpenTSDB 服务器。在多线程并发测试时,HBase HMaster 节点(S1)可以根据用户的请求自动将其分配到某个 RegionServer 节点。由于 OpenTSDB 不支持自动负载均衡,因此通过程序以平均分配的方式将查询和存储请求分发给不同的 OpenTSDB 节点。

4.2 实验结果分析

首先,模拟多台仪器同时入库操作。测试程序分别开启 1、10、50、100、200、250 个并发线程,模拟多台仪器执行插入操作,分别连续运行 3min,计算插入操作的平均事务响应时间和事务吞吐量。图 1 为写操作平均响应时间,由图 1 可以看出,在并发客户数较少时,MySQL 具有明显的并发优势,响应时间明显小于 HBase 和 OpenTSDB,随着并发客户数增多,MySQL 的平均响应时间急速上升,这也是目前基于关系数据库的地震前兆处理软件程序面临的主

要问题之一。HBase 和 OpenTSDB 并未因并发客户端的增多而大幅增加,表现出较好的可扩展性。图 2 为写操作事务吞吐情况,由图 2 也可以看出,随着并发用户数的增加,HBase 基本上呈线性上升趋势,表明本次测试并未达到系统并发处理能力的极限,还有很大的并发提升空间。OpenTSDB 也呈现出了上升趋势,但其并发性能明显不如 HBase。MySQL 数据库在并发数达到 100 后,事务吞吐量反而有所下降,表明了单机 MySQL 的并发规模应该在 150 个左右,用户数过多会导致系统性能下降。实验发现单机 MySQL 的并发规模与系统硬件配置有一定的关系,硬件配置越高,并发数越大。

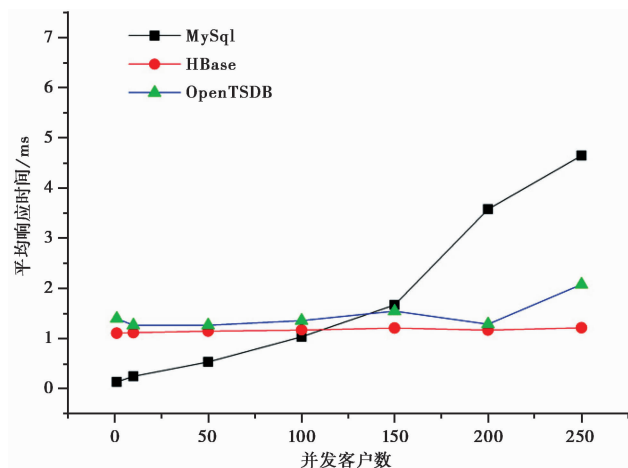


图 1 写操作平均响应时间

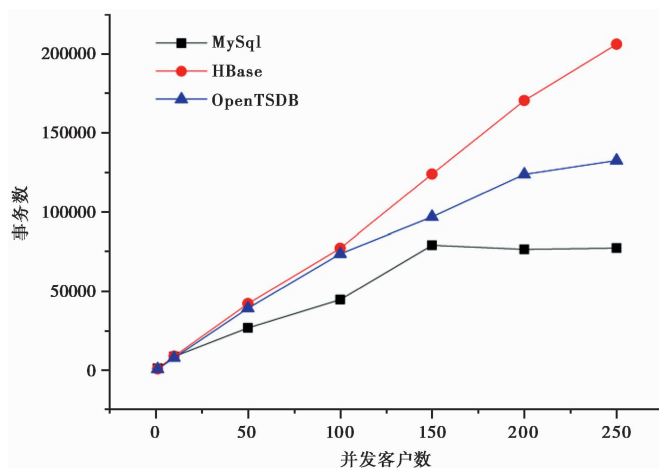


图 2 写操作事务吞吐量

根据上述业务需求分析,客户通常需要读取某台仪器某个测项在某一段时间内或多台仪器多个测项一段时间内的连续数据。在本次实验中,首先模拟产生了 1 亿条记录,然后分别测试了单用户模式和 200 个并发用户模式下获取 1 天、1 周、1 年和 10 年观测数据记录的查询操作平均响应时间。图 3 为单用户查询操作平均响应时间,图 3 表明,MySQL 和 HBase 在单用户模式下,在相同查询规模下,表现出比较稳定的查询效率,HBase 性能比 MySQL 略

好,OpenTSDB 数据库性能最差。其主要原因在于,在单用户模式下,测试程序必须先连接到 HMaster 节点,然后定位到目标数据,再连接到 HRegion 节点获取数据,而 MySQL 服务器直接在本地查询数据,故无法发挥 HBase 的并发优点。而在 OpenTSDB 方案中,由于数据存储 在 HBase 中,很有可能导致待读取数据和 OpenTSDB 服务器不在一台服务器上,查询的数据 都需要 OpenTSDB 服务器进行中转,导致耗时增多,系统性能下降明显。图 4 为 200 用户并 发查询操作平均响应时间,在图 4 中,HBase 的平均响应时间并未随着并发客户数的增多而 有明显变化,相反,MySQL 具有明显的增长现象,OpenTSDB 更是几乎直线上升,这也表明 MySQL 的单机并发性具有一定局限性。虽然 OpenTSDB 服务器可以并发处理来自多个客户 端的查询请求,但是由于业务处理和数据通常不在一起,所以这种模式的代价较高。

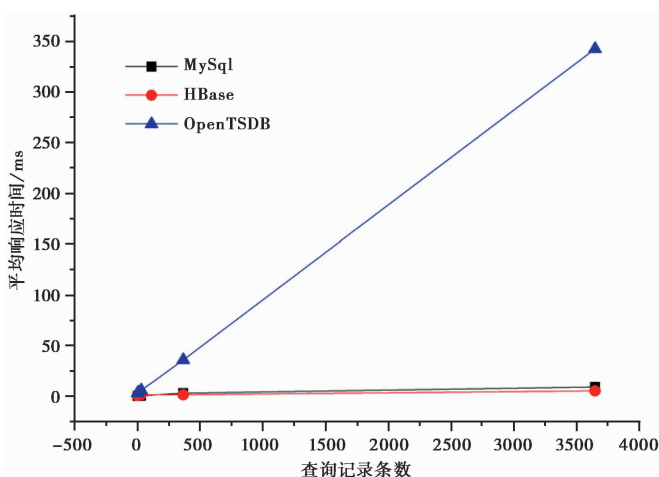


图 3 单用户查询操作平均响应时间

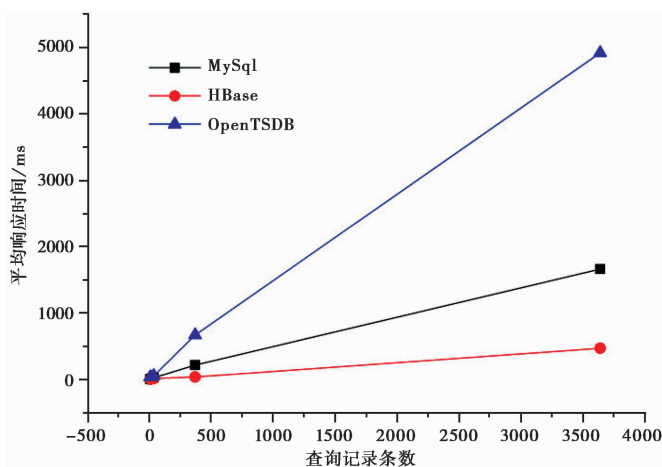


图 4 200 用户并发查询操作平均响应时间

5 结语

本文面向海量地震前兆观测数据,提出了基于 OpenTSDB 和 HBase 的地震前兆大数据

存储方案,并与传统的基于关系数据库的存储方案在读取和插入操作方面进行了对比研究。实验结果表明,基于 HBase 的存储方案无论在查询操作方面,还是在插入操作方面,均表现出了较好的性能,且具有很强的可扩展性和并发性。OpenTSDB 是一种基于 HBase 数据库,面向时间序列数据的通用数据库,虽然与 HBase 数据库一样提供了良好的可扩展性和并发性,但 OpenTSDB 并不自动提供负载均衡策略,需要用户自己在程序层自定义负载均衡策略,致使 OpenTSDB 服务处理程序和待访问数据不在同一台计算机上,固其读写性能与 HBase 方案相比略差。但由于其操作简单、接口丰富,加上良好的可扩展性和可伸缩性,因此也是大数据环境下常被采用的方案。此外,HBase 可以通过预分区,参数调优等方法进一步提高系统的性能。总之,基于云平台和大数据技术的地震大数据存储和分析业务方案将是未来地震综合数据处理的发展方向。

参考文献

- 郭凯、黄金刚、彭克银等,2017,大数据技术在海量测震数据中的研究应用,地震研究,40(2),317~323.
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,震灾防御技术,10(增刊),811~817.
刘坚、李盛乐、戴苗等,2015,基于 HBase 的地震大数据存储研究,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,软件工程,19(1),48~49,31.
周克昌、赵刚、王晨等,2013,中国地震前兆台网观测技术系统整合,中国地震,29(2),270~275.

Storage Solution of Earthquake Observation Data in Big Data Environment

Shan Weifeng^{1,2)} Teng Yuntian¹⁾ Liu Haijun²⁾ Yang Guanze²⁾

1) Institute of Geophysics, China Earthquake Administration, Beijing 100081, China

2) Institute of Disaster Prevention, Sanhe 065201, Hebei, China

Abstract The big data technology provides a new data storage solution and computing model for processing massive seismic observation data. In order to solve the problems of low read/write speed, low concurrency performance and poor scalability of existing storage schemes based on relational database, we take earthquake precursor observation data as an example, and analysis of the business requirements in detail, and then propose the two earthquake big data storage solutions based on HBase and OpenTSDB. We also construct a test platform to compare the query, insertion and concurrency performance under different storage schemes. The experimental results show that HBase and OpenTSDB-based storage schemes have better scalability and concurrency performance than relational database storage scheme, and HBase-based storage schemes have the best read and storage performance.

Key words: Big data; Earthquake; Hadoop; HBase; OpenTSDB; Concurrency