

分布式数据库 Greenplum 在地震前兆数据存储中的应用*

陈晓琳, 李盛乐, 刘 坚, 刘珠妹

(中国地震局地震研究所 地震大地测量重点实验室, 湖北 武汉 430071)

摘要: 随着地震前兆观测台网的加密、采样率的提高, 地震前兆观测数据量也在快速增加。在进行地震数据共享服务时, 需要快速获得大量数据集, 无疑对前兆共享数据库的数据处理能力提出了更高的要求。针对这一问题, 提出基于 Greenplum 数据库的地震前兆数据存储设计方案。通过搭建 Greenplum 分布式数据库环境, 实现了海量前兆数据的快速处理, 并与传统 Oracle 数据库进行对比, 结果表明: Greenplum 数据库读取前兆数据耗时更低, 对于大批量数据的读取操作, Greenplum 数据库的优势更加明显; Greenplum 数据库良好的可扩展性和对应用编程接口 (JDBC、ODBC) 的支持, 使得其在前兆数据分析处理中的应用前景广阔。

关键词: 大数据存储; 前兆数据; 分布式数据库; Greenplum 数据库

中图分类号: P315-39

文献标识码: A

文章编号: 1000-0666(2020)02-0412-05

0 引言

地震前兆观测数据作为基础数据, 在地震研究中具有广泛应用。为使地震研究人员更方便、广泛地获得全国范围内的地震前兆观测数据, 中国地震台网中心联合中国地震局地震研究所开发了地震数据对内共享服务平台 (谢有顺等, 2017)。该平台汇集了 800 多个前兆观测台站、2 900 多套前兆仪器的实时产出数据, 学科预处理、专业处理的产品数据共计 973 GB; 主要采用 B/S 架构, 用户直接通过浏览器访问位于服务器端的 Oracle 前兆共享数据库。在平台测试运行阶段, 由于用户查询、下载数据量较大, Oracle 数据库的查询、读取速度极大地制约了共享平台的访问性能。因此, 为了解决该问题, 需要研究地震前兆数据在大数据平台下的存储方案 (刘坚等, 2015)。在大数据环境下发展起来的新型分布式数据库 Greenplum 具有专为大规模查询、下载设计的并行处理架构 (Bani, Girsang, 2018); 并且拥有良好的线性扩展能力, 能方便地增加系统的存储容量和处

理能力, 这些优势都较好地契合了前兆数据共享服务的存储需求。

由于 Greenplum 数据库的良好性能, 在国内外大型企业中得到了广泛的关注。纽约证券交易所 (NYSE) 采用 Greenplum 数据库支持其海量数据的高速查询 (Kaur *et al*, 2012); 国内最早开始使用 Greenplum 数据库的是阿里巴巴 (中国) 网络技术有限公司, 其在 2008 年引入该技术, 将原有的 Oracle 数据库迁移到 Greenplum 数据库上作为数据中心, 进行产品关联分析 (何勇, 陈晓峰, 2014); 银行等金融机构采用了 Greenplum 数据库分布式架构技术来支持其大批量数据的查询需求 (张文升, 2017); 移动通信运营商将 Greenplum 数据库作为数据仓库, 利用分布式算法挖掘出隐藏在数据中的关联关系, 从而帮助运营商对客户进行精准营销 (余彪, 2015)。但值得注意的是, 以上应用均是定位于 OLAP (联机分析处理, On-Line Analytical Processing) 场景, 这是由于 Greenplum 数据库的并行处理架构, 更适合作为数据仓库使用。Greenplum 数据库不能替代前兆业务的所有场景。传统 Oracle 数据库适合前兆业务前端 OLTP (联机

* 收稿日期: 2019-02-20.

基金项目: 国家科技支撑计划 (503130108) 资助.

事务处理过程, On - Line Transaction Processing) 场景, 比如高频次的数据采集与交换的增删改操作; 而 Greenplum 数据库中基于数据仓库的信息处理过程, 更适合作为数据共享查询、下载服务的后端数据库。

本文根据前兆数据共享需求, 制定了前兆数据在 Greenplum 数据库下的存储策略。基于地震行业网搭建了 Greenplum 分布式数据库环境, 并从前兆数据读取性能、数据库可扩展性能、前兆应用软件兼容性等方面对 Greenplum 数据库与 Oracle 数据库进行了对比测试。

1 Greenplum 数据库原理

目前, 大数据存储或计算是热点问题, 涌现了许多的大数据存储技术。而 Greenplum 数据库作为关系型数据库, 更容易和目前使用的 Oracle 数据库进行对接、转换。另外 Greenplum 数据库完全支持 SQL 标准, 支持 ODBC 和 JDBC, 从应用编程接口上讲, 更利于原有的地震数据分析应用软件迁移到大数据平台。

1.1 分布式并行架构

Greenplum 数据库采用了大规模并行处理架构 (Massively Parallel Processing, 以下简称 MPP), MPP 是由 2 个或多个处理器构成的系统。每个处理器都有自己的内存、操作系统和磁盘, 它们并行执行用户操作 (陈达伦等, 2016)。主节点 Master 是 Greenplum 数据库系统的入口点 (图 1), 是与客户端连接并提交 SQL 语句的数据库实例。

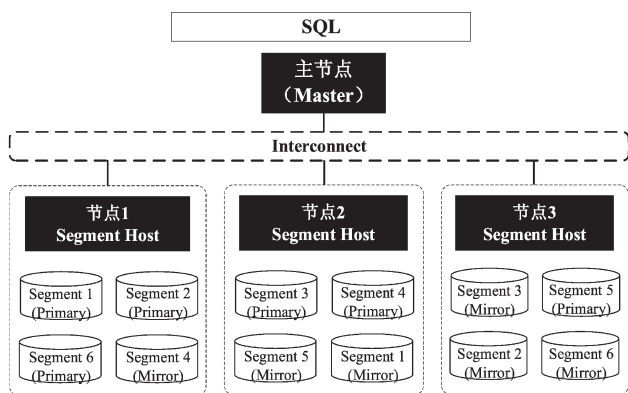


图 1 Greenplum 数据库集群架构

Fig. 1 Greenplum database cluster framework

主节点 Master 将工作负载分发给其它数据库实例 (Segment Instance) 进行存储和处理数据。Interconnect (网络层) 支持在不同的 Segment 实例 (独立的 PostgreSQL 数据库) 之间进行通信, 并允许系统作为一个逻辑数据库运行。由于 Greenplum 数据库使用这种高性能的系统架构来分担数据仓库的负载, 并可以并行使用系统的所有资源来处理用户任务。因此, Greenplum 数据库在大数据量的查询、下载时表现出了优越的性能。

Greenplum 数据库是基于 PostgreSQL 的数据库集群 (Waas, 2009), 它们共同工作以呈现单个数据库映像。大多数情况下, 在 SQL 语句、特性、配置选项和最终用户功能方面, Greenplum 数据库与 PostgreSQL 数据库非常相似, 数据库用户与 Greenplum 数据库交互, 就像与普通的 PostgreSQL 数据库交互一样。但是 Greenplum 数据库对 PostgreSQL 数据库的内部结构进行了修改或补充, 以支持它的并行结构。例如, 系统目录、优化器、查询器和事务管理器组件已经被修改和增强, 以便能够让所有分布式 PostgreSQL 数据库同时并发执行。

1.2 数据分布与同步

由于 Greenplum 数据库的分布式架构, 因此所有的表都会被切片成若干数据片存储在 Segment 上。分区策略可以为随机分布和 Hash 分布。随机分布中数据被随机分散到每个 Segment, 虽然这种分布方式可以保证数据被均匀分布到每个 Segment, 但是在进行表的关联操作时, 需要根据关联键重分布数据, 所以随机分布性能较差。Hash 分布可以指定表的一列或者多列组合作为 Hash Key 来进行数据分布。Greenplum 数据库会根据指定的 Hash Key 列来计算每一行数据对应的 Hash 值, 并映射至相应的 Segment 实例。对于 Greenplum 数据库, 如果用户创建表时未指定数据分布方式, 则默认为 Hash 分布, 并且将主键作为 Hash Key, 如果没有设置主键则选择第一列作为 Hash Key。

为了避免单点故障, Greenplum 数据库对主节点和各数据节点均采用备份策略。当主节点 Master 发生故障时, 激活主节点备份 Standby 主机接管 Master 事务, 它们之间通过流复制技术实现同步复制。如图 1 所示, 主数据节点 (Primary) 和它的镜像 (Mirror) 是分配在不同的主机上, 以防止某

一数据节点故障导致数据库的访问异常。主节点和其数据镜像之间通过文件操作级来实现数据的同步 (GISEarth, 2016)。

2 前兆数据建库关键技术

2.1 集群环境搭建

在行业网内搭建 Greenplum 分布式数据库环境, 初始化搭建时由 1 个主节点 (Master) 和 3 个数据节点 (Segment Host) 组成, 节点间使用千兆网络连接, 节点配置见表 1。

高可用性设计: 主节点 (Master) 负责解析、分发 SQL 语句, 收集、汇总查询数据等工作。因

此必须保证主节点的高可用性, 这里同时将 Segment Host3 机器设置为主节点的备份节点 (Master Standby)。每个数据节点都配备镜像数据, 当主数据节点发生故障时, Mirror 节点会自动切换为 Primary, 不会对前台应用产生影响。

在线扩展: 各类前兆观测台站 800 多个, 前兆仪器 2 900 多套, 各测项采样率逐渐提高, 秒采样逐渐成为趋势。秒采样的数据量比分采样的数据量提高了 60 倍, 并且后期设计要接入气象观测数据。当数据量日益增大, 现有集群不能满足需求时, 可以通过增加集群数量对 Greenplum 数据库进行线性扩展。扩展过程中, 业务可以继续运行, 不需要宕机。

表 1 Greenplum 数据库集群服务器参数
Tab. 1 Greenplum database cluster server parameters

指标	Master Host	Segment Host1	Segment Host2	Segment Host3
操作系统	CentOS 6.4 64 位	CentOS 6.4 64 位	CentOS 6.4 64 位	CentOS 6.4 64 位
CPU	Intel(R) Xeon(R)	Intel(R) Xeon(R)	Intel(R) Xeon(R)	Intel(R) Xeon(R)
	CPU E5 - 2650 v2	CPU E5 - 2650 v2	CPU E5 - 2650 v2	CPU E5 - 2650 v2
内存	128 G	128 G	128 G	128 G
带宽	1 000 Mbps	1 000 Mbps	1 000 Mbps	1 000 Mbps
Segment 数量	0	3	3	3
版本	greenplum - db - 4.3.8.2 - build - 1 - RHEL5 - x86_64			

2.2 存储分布策略

对于 Greenplum 数据库, 影响 SQL 查询性能的最重要因素是数据分布是否均匀。如果数据在 Segment 分布不均匀, 会导致短板效应, 因此设计合适的存储分布策略至关重要。一般情况下, 如果数据经常被高并发的键值或离散查询, 将查询条件的列作为分布列, 这样不需要连接到所有的 Segment 去查, 可以大大提高并发能力。因此根据前兆数据共享需求, 制定了如下存储规则: 指定表的主键 (台站编码、测点编码、测项编码、开始时间) 作为分布键。

在 Greenplum 数据库中除了将数据分布到各 Segment 外, 还要对各 Segment 上的数据进行分区, 分布和分区共同决定数据的查询效率。分区是从逻辑上对同一个 Segment 上的表数据进行划分, 这样可以优化查询性能, 不会影响数据在各个 Segment 上的分布情况。目前 Greenplum 数据库支持 LIST 和 RANGE 两种分区类型。分区的目的是尽可

能地缩小 QUERY 需要扫描的数据量, 因此必须和查询条件相关联。鉴于此, 本文以时间作为分区键, 选择 RANGE 分区类型。

2.3 字段类型设计

Greenplum 数据库的数据类型十分丰富, 基本与 Oracle 数据库的一样, 仅在少数数据类型上有区别 (Rajput *et al*, 2013)。在建设前兆 Greenplum 数据库时, 参照“十五”前兆数据库 (周克昌等, 2010) 中的常用数据类型选择了对应的数据类型, 见表 2。

表 2 Greenplum 数据库字段类型
Tab. 2 Greenplum database field type

Oracle	Greenplum	字段说明
date	date	日期和时间
char	char	字符 (定长)
varchar2	varchar2	字符 (变长)
number	numeric	数值 (声明精度)
clob	text	文本数据
blob	bytea	二进制数据

3 性能测试与结果分析

为了更好地进行前兆共享数据库的选型，本文从数据查询、集群扩展等方面对 Greenplum 数据库进行了测试实验，并给出了测试结果。

表3 查询测试结果

Tab. 3 Query test results

测试数据	测试项目	日期范围	记录数	Oracle	Greenplum
邢台隆尧红山台	小数据量查询测试	2019-01-01—2019-01-07	63 条	2 s	1.9 s
秒钟值数据	大数据量查询测试	2017-01-01—2019-01-01	5 902 条	278 s	21 s

从表3测试结果来看，对于小数据量的查询处理，Oracle 和 Greenplum 数据库均可以在较短的时间内完成处理，用户等待时间差距不明显；而对于大批量数据的查询处理，Oracle 数据库等待时间较长，Greenplum 数据库则在很短的时间即可完成查询任务。显然，对于大批量数据的查询处理，Greenplum 数据库明显优于 Oracle 数据库。这是由于 Greenplum 数据库将查询任务分布至各个数据节点进行并行处理，充分利用了各数据节点的 CPU 和 IO 能力，使数据分析处理能力得到提升。

3.2 可扩展性能测试

Greenplum 数据库的重要特性是能在数据规模增大时便捷地进行线性扩展，通过增加节点数量高效提升数据库的性能。为此，本文在基础集群为4节点的基础上，逐步将集群增加至6，8节点，对数据库的扩展性进行了测试，结果见表4。

表4 不同节点集群上查询测试结果

Tab. 4 Query test results on different node clusters

数据库类型	节点个数	查询时间/s
Oracle	—	278
	4	21
Greenplum	6	11
	8	8

测试结果显示，随着节点数的增加，Greenplum 数据库的查询效率也随之提升。这种良好的横向扩展性可以支持数据库良好的成长，以应对不断增长的前兆观测数据。

3.1 数据读取性能测试

以 QZ_312_DYS_02（地磁变化记录秒数据）表为例，对部署于相同网段的 Oracle 和 Greenplum 数据库进行数据读取速度测试。测试程序采用 Java 编写，利用 JDBC 驱动远程连接数据库。分别进行了小数据量和大数据量的查询测试，查询结果见表3。

3.3 应用兼容性测试

现有的前兆各学科各种数据处理软件基于 Oracle 数据库运行多年（王建军等，2019），其主要数据库操作部分均为查询读取，较好地匹配了 Greenplum 数据库的数据读取优势。如果能将其后台数据库迁移至 Greenplum 数据库，将会极大地提升数据处理效率。而且本文所采用的 Greenplum 数据库完全支持 SQL 标准，支持 ODBC 和 JDBC。因此只需修改这些软件的数据库接口部分的代码，便可以将底层数据库经由 Oracle 转换为 Greenplum。以前兆数据处理中的周期分析和前兆数据共享网站为例，分别测试 Greenplum 数据库对 SQL 标准及 ODBC 和 JDBC 的支持。

表5 Greenplum 数据库的 ODBC 及 JDBC 访问测试

Tab. 5 ODBC and JDBC access tests for the Greenplum database

测试项目	ODBC	JDBC
测试环境	VS2013	Eclipse
示例应用	前兆数据周期分析	前兆数据共享网站
连接驱动	Npgsql.dll,	postgresql-8.2-507.
	Mono.Security.dll	jdbc3.jar
主要类库	NpgsqlConnection,	DriverManager,
	NpgsqlDataAdapter	Connection, Statement

4 结论

本文提出将分布式数据库 Greenplum 用于前兆数据共享服务，以满足大批量前兆数据的快速读取需求。通过在地震行业网内搭建分布式集群服务器，部署了基于 Greenplum 的前兆共享数据库。

性能测试实验表明,与传统 Oracle 数据库相比,基于 Greenplum 的前兆共享数据库在大批量的数据查询处理上,具有显著的优势,并且 Greenplum 数据库架构具有良好的可扩展性,随着节点数的增加,查询效率能得到进一步提升。Greenplum 数据库作为大数据环境下发展起来的数据库,增加了可编程并行分析挖掘功能,并提供广泛的语言支持,使得现有的前兆分析应用程序移植起来极为方便。因此,利用 Greenplum 数据库开展前兆数据库分析挖掘是下一步的研究方向。

测试环境搭建过程中,得到南京云创存储科技有限公司地震研发组成员马鸣、陈旭等的大力支持和热心帮助,在此表示感谢!

参考文献:

- 陈达伦,陈荣国,谢炯. 2016. 基于 MPP 架构的并行空间数据库原型系统的设计与实现[J]. 地球信息科学学报,18(2):151-159.
- GISEarth. 2016. 开源大数据引擎:Greenplum 数据库架构分析[EB/OL]. (2016-06-29)[2019-01-15]. <https://blog.csdn.net/happyduoduo1/article/details/51784874>.
- 何勇,陈晓峰. 2014. Greenplum 企业应用实战[M]. 北京:机械工业出版社.
- 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.
- 王建军,赵银刚,刘高川. 2019. 地震前兆 Oracle LOB 数据压缩与交换及其访问效率研究[J]. 地震研究,42(3):447-453.
- 谢有顺,李盛乐,刘小利,等. 2017. 面向地震行业的地震数据共享服务平台的设计与实现[J]. 大地测量与地球动力学,37(5):546-550.
- 余彪. 2015. 基于 Greenplum 的移动通信数据业务分析[D]. 长春:吉林大学.
- 张文升. 2017. 分布式数据库 Greenplum 研究与应用[J]. 金融科技时代,(6):44-46.
- 周克昌,蒋春花,纪寿文,等. 2010. 地震前兆数据库系统设计[J]. 地震,30(2):143-151.
- Bani F C D, Girsang A S. 2018. Implementation of database massively parallel processing system to build scalability on process data warehouse[J]. Procedia Computer Science,135:68-79.
- Kaur H, Agrawal P, Dhiman A. 2012. Visualizing clouds on different stages of DWH - an introduction to data warehouse as a service[C]// Computing Sciences (ICCS), 2012 International Conference on IEEE.
- Rajput E, Yadav H, Singh A. 2013. Comparative study of EMC greenplum and Oracle exadata[J]. Journal of Engineering Computers & Applied Sciences,2(4):2319-5606.
- Waas F M. 2009. Beyond conventional data warehousing—massively parallel data processing with Greenplum database[M]. Berlin:Springer,89-96.

Application of Distributed Database Greenplum in Earthquake Precursor Data Storage

CHEN Xiaolin, LI Shengle, LIU Jian, LIU Zhumei

(Key Laboratory of Earthquake Geodesy, Institute of Seismology, China Earthquake Administration, Wuhan 430071, Hubei, China)

Abstract

With the encryption of precursor observation networks and the improvement of sampling rate, the quantity of earthquake precursor observation data increases rapidly. It is a higher requirement for the data processing ability of the precursor database in the seismic data sharing service that how to make users quickly obtain large data sets. Aiming at this problem, the Greenplum-based precursor data storage method is proposed. Through setting up the Greenplum distributed database environment, the rapid processing of massive precursor data is realized. Performance testing experiments of precursor data show that the Greenplum database takes less time to read the precursor data, and the advantages of Greenplum database are more obvious for large-volume data operations, compared with the traditional Oracle11g database. In addition, the good extensibility of Greenplum database and the support of application interface (JDBC, ODBC) make its promising application in precursor data analysis and processing.

Keywords: big data storage; precursor data; distributed database; Greenplum database