地震观测集成数据库的优化研究*

丁 睿

（上海市地震局，上海 200062）

摘要：利用硬件设备、网络、存储、系统架构等高可用环境升级的技术手段，并采用 Oracle 数据库参数调整、索引添加、分区等方法对集成数据库进行性能调优。测试结果表明，优化后的集成观测数据库性能得到有效的提升。同时，通过 SQL 语句解析的深入分析，对业务应用软件的设计开发提出 SQL 语句优化和实时波形存储方式改变的修改建议。

关键词：地震观测集成数据库；高可用环境；数据库优化；SQL 优化

中图分类号：P315.3　　**文献标识码**：A　　**文章编号**：1000－0666(2014)04－0654－06

0 引言

在地震数据共享和产品化服务的背景下，地震信息资源整合成为一种必然趋势。笔者通过同构数据迁移和异构数据整合，将泰德测震、强震系统，JOPENS 测震系统和前兆数据处理系统在数据存储层进行整合，形成了地震观测集成数据库。业务数据集中管理的模式有效利用资源，提高了工作效率（丁睿，秦浩文，2011）。但运行数年来，观测集成数据库存在多业务用户资源抢占，系统高峰时运行速度降低，数据坏块出现频率增高等问题，随着观测数据的增加，核心数据库的负载增强，数据库的响应时间明显下降。因此，对于地震观测集成数据库的性能优化十分迫切。

1 数据库优化的目标和策略

数据库的性能调整与优化是一项活动，它通过优化应用程序、修改系统参数、改变系统配置来提高系统性能（刘博，2007）。影响数据库服务系统性能的因素很多，一般来说，数据库优化包括对硬件、操作系统与数据库管理系统的调整，以及对访问这些组件的业务应用详细分析和优化。评价一套数据库运行效能，一般以系统吞吐量、系统响应时间、支持用户能力、系统容错能力和数据加载时间这 5 个方面为准则（刘玉强，2007），这也是数据库优化需要达到的目标。

数据库性能的调整贯穿于系统的设计、开发、调试和运行等阶段，涉及面广，对 Oracle 数据库进行性能优化时，应当按照一定的顺序进行（林键，2012）。在应用系统的设计、开发阶段，对其数据库逻辑结构和物理结构进行优化设计，使之在满足需求条件的情况下，系统性能达到最佳，系统开销达到最小，可以避免正式上线后的一些不必要调整或者代价很大的调整（邓明元，屈辉立，2006）。在数据库运行阶段，多采用操作系统级、数据库级的一些优化措施来使得性能最佳。

笔者以上海地震观测集成数据库为优化对象。优化前存在多用户抢占系统资源、数据库负载过高，业务高峰期系统响应缓慢，核心数据安全保护较弱，容易出现坏块等问题。由于该集成数据库为运行中的系统，对其进行的性能优化调整主要集中在高可用环境和 Oracle 数据库上。同时，通过分析 SQL 语句，对业务应用软件的设计开发提出修改建议。

2 集成数据库的性能调优

针对上海地震观测集成数据库出现的性能瓶颈和运行特性，需要对两方面内容进行多层级、全方位的性能调整和优化：一是数据服务系统依托的高可用环境，包括服务器、网络、存储、数据库软件和系统架构等；二是 Oracle 数据库，包括内存参数优化、索引优化、分区优化等。

* **收稿日期**：2013－12－25.

基金项目：地震科技星火计划项目—地震业务集成数据库自动监控、管理和优化研究（XH12018Y）资助.

2.1　高可用环境升级

高可用性有广义和狭义之分。本文的研究对象是广义高可用环境，指整个系统的高可用，包括系统失败或崩溃，应用层或中间层错误，网络失败，存储介质失败，人为失误、分级与容灾，计划宕机与维护（陈吉平，2008）。数字地震网络项目竣工后两年内，通过数据整合，逐渐形成了由 SUN Cluster 的小型机服务器、Oracle10gR2 数据库管理软件、Solaris Cluster HA 操作系统，备份服务器（运行 EMC legato 备份软件）、TB 级的 IPSAN 存储、以及三层双交换机组成的一套上海地震观测数据服务系统。

随着观测节点和实时数据量的增加，业务系统对数据库的访问频率增强，观测集成数据库的压力越来越大。业务高峰期，使用 mpstat、echo::memstat | mdb -k 等命令对系统负载进行监控，可以看到系统资源损耗严重。

高可用环境升级采取由分到总的策略，对服务器硬件、网络、存储和数据库体系架构分别进行分析测试，再综合择优，确定最合适的高可用环境升级方案。

（1）数据库服务器

数据库服务器硬件选择在传统机架式 x86 服务器、小型机和刀片服务器之间进行，比较了 DELL、IBM、Oracle、Cisco 等品牌。考虑到数据中心的统一资源管理，业务快速部署和低能耗需求，选择 Cisco UCS（统一计算系统）为平台，将核心数据库服务器从 Sun Fire V490 更换为 Cisco UCS Blade Server B200 M2 刀片服务器。CPU 由原来的 4C 1.05 GHz 提升到 16C 2.67 GHz，内存由原来 8 GB 提升到 32 GB。当多用户进行各种业务操作时，每一个用户的操作请求，CPU 都会分配一个线程来处理。CPU 核数的增加，使 CPU 总线程数也得到了增加，满足同一时间多用户进行业务操作。CPU 主频的增大可以进一步提高 CPU 处理用户请求的速度。

（2）网络

Cisco UCS 的统一矩阵 Unified Fabric，将服务器的所有网络流量通过统一的 fabric interconnects 传输，进行统一处理和管理，大大减少了网络适配器，刀片服务器交换设备和网络布线，最终提高整体性能。局域网由千兆网络变更为万兆网络，带宽的增加，使数据传输更快速，对业务应用响应也起到一定的提升作用。

（3）存储

存储的选择在 IP SAN（以太网存储）和 FC SAN（光纤存储）之间。高可用环境升级中，核心存储由原来的 H3C IX1000 IP SAN 更换为 HDS VSP FC SAN，即将过去基于 IP 网络的存储系统变更为采用光纤通道的存储设备。HDS VSP FC SAN 采用 4 GB/s 的光纤通道和 512 GB 数据缓存。FC 由于协议设计的特点使得其具有先天的安全特性，与 LAN 业务网络相隔离，使存储与服务器之间进行数据传输时更快捷稳定，存储数据流也不会占用业务网络带宽。SAS 盘比 SATA 盘在磁盘读写性能上也有一定提高。存储的升级对数据库服务器物理读取或写入存储中数据的性能有一定提升。

（4）数据库软件

集成后的观测业务数据库采用 Oracle 10g R2 作为数据库管理系统，软件版本由原来的 10.2.0.3 升级到 10.2.0.4，更稳定可靠。

（5）数据库体系架构

业务集成数据库由 HA 方式的单实例变更为双节点 RAC。比起单实例数据库，RAC 数据库具有负载均衡的特性，Cache Fusion 可以解决共享缓存（Cache）并发问题（刘增军，2006）。升级后的业务观测集成数据库将文件存储改为 ASM 存储方式，具有可扩展性，当磁盘满后，可以手动在线添加。ASM 磁盘组中的磁盘做成镜像冗余方式，提高可靠性。ASM 对从磁盘中进行数据读取的性能有一定提升作用。

升级后上海地震观测集成数据库高可用环境如图 1 所示。

2.2　数据库优化

对观测集成 Oracle 数据库采用调整内存参数、增加索引和对大数据表进行分区 3 种方式进行优化。

（1）内存参数调整

内存参数的调整主要是指 Oracle 数据库的系统全局区 SGA（System Global Area）的调整。SGA 是 Oracle 数据库的心脏，是对数据库数据进行快速访问的一个系统区域，可以被服务器和用户共享（肖军，2004）。SGA 主要由 3 部分构成：共享池（Share Pool）、数据缓冲区（Data Buffers）和日志缓冲区（Redo Log Buffers）。程序全局区 PGA（Program Global Area）也是一个内存区域，包含了

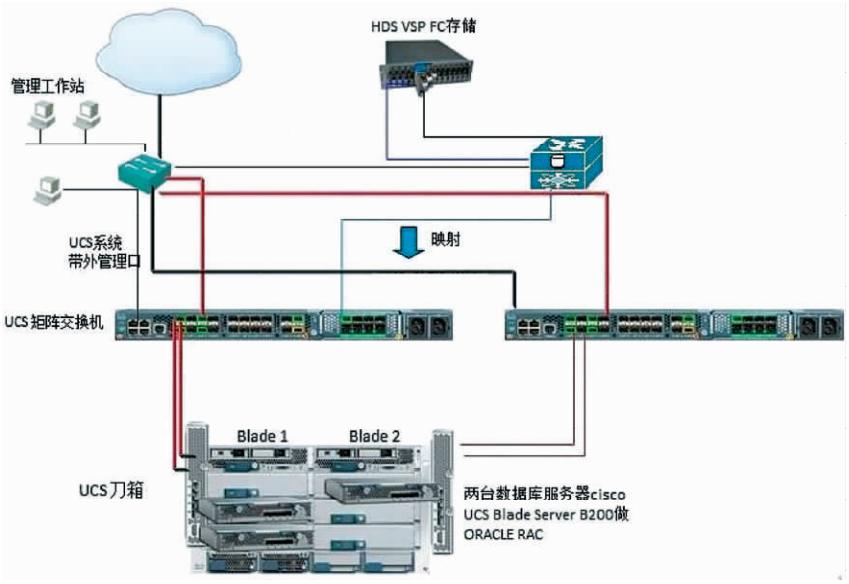

图 1 升级后的地震观测集成数据库高可用环境

Fig. 1 The high availability environment of seismic observation integrated database after the upgrade

与某个特定服务器进程相关的数据和控制信息。每个进程都具有自己私有的 PGA 区。

通过查询 v $ systat 视图中的各种指标和在业务高峰期收集的 AWR 报告，发现系统运行速度过慢的主要原因是 SGA 内部的数据缓冲区命中率过低，数据库系统在查询时不能从共享池或者数据缓冲区内找到已经缓存的数据，从而频繁地到磁盘上读取数据，消耗系统资源，造成等待时间过长。

通过多方测试，观测集成数据库中将 sga_ target 参数从原来的 1.5 GB 扩大到 14 GB，pga_ aggregate_ target 从 809 MB 扩大到 4 805 MB。调整内存参数后，在业务高峰期收集 Oracle RAC 两个节点的 49minsAWR 报告，从中可以看到测试时间段一节点数据库内存分配为：buffer cache 8 176 MB，shared pool 5 248 MB，PGA 4 805 MB。测试时间段二节点数据库内存分配为：buffer cache 6 784 MB，shared pool size 6 656 MB，PGA 4 805 MB。根据各自节点的 AWR 报告中的 Buffer Pool Advisory 和 PGA Memory Advisory，可以看出两节点 buffer cache 值所在的因子下，比原来 buffer cache 值所在的因子下产生更低的额外物理读；两节点 PGA 的值对应各自因子下，比原来 PGA 的值对应的因子下产生更低的额外物理写。同时从各自节点上 AWR 报告中可以看出两个节点的 Buffer Hit % 都在 100%，数据块在数据缓冲区中命中率正常。从 TOP 5 等待时间中没有发现明显耗资源的等待事件。在测试过程中各节点的 AWR 报告中 DB Time 都远小于内核数 * Elapsed，说明在测试过程中数据库压力不大。扩大 SGA 和 PGA 对数据库性能有一定提升。

另外，从 AWR 报告上还发现，数据库的软解析值很高，但 Execute to Parse% 比例却很低。查看 session_ cached_ cursors 的使用情况（表 1），发现其使用率为 100%，这个参数限制了在 PGA 内 session cursor cache list 的长度，提供快速软分析的功能，比软解析性能更高（谭磊，2010）。因此，需要根据应用情况适当增大这个参数值，session_ cached_ cursors 数量要小于 open_ cursors，采用下面的命令将 session_ cached_ cursors 值增加到 100：alter system set session_ cached_ cursors = 100 scope = spfile。

表 1 session_ cached_ cursors **使用情况**

Tab. 1 Service condition of session_ cached_ cursors

参数	数值	使用率
session_ cached_ cursors	20	100%
open_ cursors	3000	1%

（2）添加索引

对 JOPENS 系统的 MySQL 数据库进行异构整合，迁移到地震观测集成 Oracle 数据库中，并为其创建 MATE 用户。为了提高测震业务用户读取波形数据的速度，分别对 MATE 用户下的 WAVE-

FORM_ CON 表中 CHANNEL_ ID 和 START_ TIME 字段添加索引，索引名为 WFC_ CHANNEL_ IDX 和 WFE _ STARTTIME _ IDX。添加索引后，当 JOPENS 测震应用进行读取某一频道的波形数据，或者读取某一时间点（时间段）的波形数据时，数据库通过索引扫描查询数据，避免顺序扫描整个表，从而提高应用查询效率。

（3）使用分区表存储波形数据

Oraclel 的分区（Partitfon）技术，是为解决数据库中大表或者索引读写速度过慢而提出的解决方案。通过把大表和大索引拆分成更小的块（称为分区），与原来相比，因为分区变小，所以系统访问分区的速度更快、效率也更高，但分区的存在对用户而言是透明的。系统可单独访问这些小尺寸的表或索引，也可以以组或整体方式访问（姜维智，2005）。

泰德测震、强震的表空间 seism，每个月的数据增长有 300 ~400 GB。当前泰德应用的测震、强震表都在 seism 表空间中。该业务应用数据量大，实时波形数据不断增长，并且应用经常会做查询某一时间点的测震或者强震少量的数据。采用在线重定义的方法对后续收集的数据使用分区技术，对测震、强震分别按时间范围（按月）建立分区表，同时对站点、频道和起始时间字段建立分区索引。分区表的使用可以增强维护性，当表某个分区出现故障，只需修复该分区，不需要对整表进行修复；可以均衡 I/O，把不同的分区映射到磁盘以平衡 I/O；可以改善查询性能，对分区对象的查询可以仅搜索自己关心的分区，排除其它不相干的数据，从而提高检索速度，改善数据库的性能。

2.3　整体优化后的性能对比

（1）系统资源状况

经过整体优化后，观测集成数据库系统资源充足，可查看 CPU 和内存使用，如下所示：

[root@ DZJ1 ~] # mpstat

Linux 2.6.18 -274.el5 (DZJ1) 03/25/2013

03：33：59 PM CPU % user % nice % sys % iowait % irq % soft % steal % idle intr/s

03：33：59 PM all 5.05 0.00 0.43 0.20 0.02 0.26 0.00 94.03 765.19

[root@ DZJ1 ~] # free -m

total used free shared buffers cached

Mem：32112 18103 14009 0 301 15872

-/+ buffers/cache：1929 30183

Swap：34111 0 34111

（2）数据库负载状况

在测震、强震、前兆等业务高峰期，观测集成数据库两个节点负载明显降低，整体性能如图 2 所示：

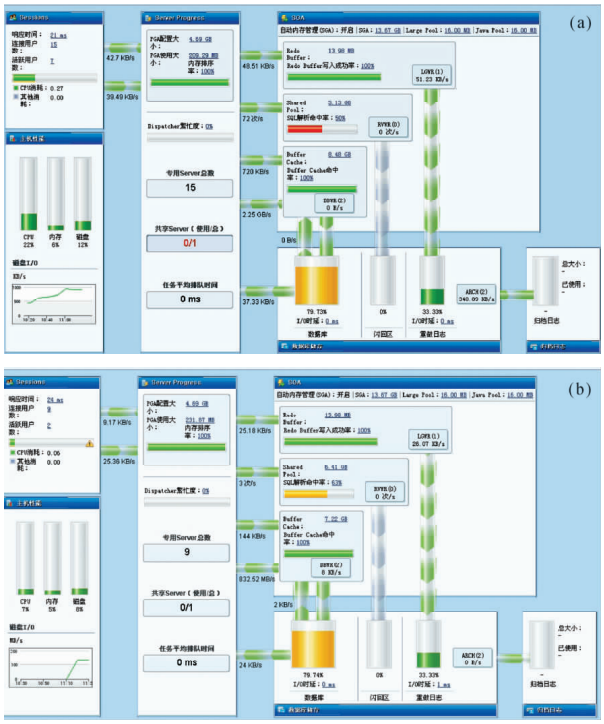

图 2　优化后的集成数据库节点 1（a）、节点 2（b）性能图

Fig. 2　Performance of Nade 1 (a), Nade 2 (b) of integrated database after optimization

（3）高峰期业务系统响应时间对比

在业务正常运行情况下，抽选对用户影响较大的几个业务场景，执行相关 SQL 语句，测试数据库耗时，对比如表 1 所示。

根据测试对比结果，地震观测集成数据库高可用环境升级后，使系统整体（包括数据库服务器、存储、网络）性能得到了一定提升，系统硬件资源已不存在明显性能瓶颈，同时 Oracle 数据库软件版本和系统架构也进行了升级和调整，降低数据库负载，更加符合业务应用的需求。对 Oracle 数据库的优化（包括内存参数、索引和分区优化）后，集成数据库的性能有了一定的提升，业务高峰时期数据库压力不大，原来响应缓慢应用

操作在整体优化后执行效率显著的提升。独立的核心数据存储区将业务观测数据与一般数据隔离，保障了安全性。另外，整体优化后，集成数据库再没有出现坏块。

表 2 数据库优化前后运行时间对比

Tab. 2 Run time contrast before and after Database optimization

业务场景	读取 59 个强震台 200 Hz 一小时内的波形数据	读取 10 个测震台 100 Hz 一天内的波形数据	查看核心数据库中所有站点信息	核心数据库监控信息查询	前兆用户查询一个台站的一种观测手段
数据库优化前运行时间	19 min 12 s	23 min 30 s	5 min 24 s	48 s	18 s
数据库优化后运行时间	2 min 10 s	3 min 5 s	6 s	3 s	1 s
减少时间	17 min 2 s	20 min 25 s	5 min 18 s	45 s	17 s

3 需要进一步进行优化研究的问题

经过高可用环境升级和 Oracle 数据库优化，上海地震观测集成数据库得到较大的提升。但本次性能调优是在数据库、业务应用系统运行阶段进行，设计、开发阶段的一些影响性能的问题，还需要进行深度优化研究。

（1）应用 SQL 语句优化

通过数据库监控设备查询目前正在运行的观测集成数据库一周的 SQL 解析，发现命中率偏低，平均未命中率达到 80%（20% 以内较正常），而 SQL 硬解析比例偏高，平均比例达到 25% 以上（10% 以内较正常）。从监控设备抓取到的业务应用 SQL 语句可看出，存在大量硬解析比例较高的查询或删除语句。根据监测结果，需要对测震、前兆等观测系统应用中的部分 SQL 语句进行优化，SQL 语句的优化需要业务软件开发工程师在应用软件层面进行。

（2）实时波形文件存储方式改变

目前测震、强震系统中的波形数据存放在波形表的 BLOB 字段中。对大数据集，存储二进制数据将会使数据库文件迅速变大，难以控制数据库的大小。建议波形数据要存放到文件系统中，数据库中存放该波形文件相应的链接（如文件路径和名称）。需要时应用程序中通过链接来取得文件中的数据。这样可以节省数据空间，避免了数据库过分膨胀，同时将二进制数据存储到文件系统中也会获得较快的读取效率。

4 结论

通过对高可用环境中的硬件、软件、系统架构升级和对 oracle 数据库进行内存参数调整、添加索引及采用分区技术优化，地震观测集成数据库的性能得到较明显的提高。然而数据库性能优化是一个复杂且需反复进行的工作，需要在大量的实践工作中不断地积累经验，进一步改善集成数据库的服务效能。

本文在撰写过程中得到上海市地震局信息网络室同志，中软公司、泰德公司技术工程师的帮助，在此表示衷心感谢。

参考文献：

陈吉平 . 2008. ORACLE 高可用环境——企业级高可用数据库架构、实战与经验总结[M]. 北京：电子工业出版社，40 – 41.

邓明元，屈辉立 . 2006. 局域网的 Oracle 数据库管理[J]. 电脑与信息技术，14(1)：54 – 58.

丁睿，秦浩文 . 2011. Oracle 高可用环境下地震业务数据集中管理探索[J]. 地震地磁观测与研究，32(2)：75 – 80.

姜维智 . 2005. ORACLE 数据库性能优化[D]. 成都：电子科技大学，40 – 41.

林键 . 2012. 基础地理空间数据库性能优化的研究[J]. 浙江测绘，(4)：46 – 48.

刘博 . 2007. Oracle 数据库性能调整与优化[D]. 大连：大连理工大学，10 – 11.

刘玉强 . 2007. 基于 ORACLE(OLTP)数据库性能优化方案的研究与实施[D]. 北京：北京邮电大学，5 – 7.

刘增军 . 2006. 高可用性数据库系统研究、应用与性能优化[D]. 长沙：国防科技大学大学，14 – 17.

谭磊 . 2010. 基于等待事件的 Oracle 数据库调优与实时监控研究[D]. 成都：成都理工大学，15 – 16.

肖军 . 2004. ORACLE 数据库性能调整与优化[D]. 武汉：武汉大学，12 – 13.

Research on Optimization of Earthquake Observation Integrated Database

DING Rui

(*Earthquake Administration of Shanghai Municipality*, *Shanghai* 200062, *China*)

Abstract

Using technical means of available service environment upgrading, such as the hardware equipment, network, storage, system architecture etc., we do the performance tuning on the integrated database by the methods of adjustment, adding index and partition etc. of Oracle database parameter. The test result show that the performance of integrated observation database improved efficiently after optimization. Through the in-depth analysis of SQL statement parsing, we put forward modification proposals to optimize SQL statements and change waveform storage method in real time for the design and development of business application software

Key words: earthquake observation integrated database; available service environment; database optimization; SQL optimization