

通过 C++ 类封装 Oracle 调用接口实现 地热前兆数据库的快速建库与访问^{*}

王 军¹, 赵 刚¹, 何案华¹, 郭藐西¹, 郭柏林¹, 秦久刚²

(1. 中国地震局地壳应力研究所, 北京 100085 2 中国地震局地壳运动监测工程研究中心, 北京 100036)

摘要: 为了进一步推动地热前兆方法的发展, 需要全面系统地对地热前兆方法加以认真研究, 并应尽可能全面地搜集地热前兆观测资料, 建立起地热前兆数据库。使用 C++ 语言, 将 Oracle 调用接口封装成 QEnvironment, QStatement, QDirPathLoader 和 QOCException 四个类, 可实现地热前兆应用数据库的快速建库和访问, 同时增强代码的可读性和易维护性。

关键词: 地热前兆; 数据库; OCI; Oracle; C++; 直接路径装载

中图分类号: TP274 **文献标志码:** A **文章编号:** 1000-0666(2008)03-0284-05

0 引言

地热前兆方法是使用系统科学方法, 对稳态地热场及正常地热场的未来进行预测, 排除干扰, 研究附加地热场的形成、特征、活动过程及与地震的关系, 从而达到监测预报地震的目的 (付子忠等, 1990)。1984 年, 国家地震局地壳应力研究所地热组在云南省下关市滇西地震预报实验场开始了最早的地热前兆实验观测研究, 在地震预报实践中取得了一定成绩。“八五”期间, 在原来云南省地热前兆实验观测研究的基础上, 研究人员对地热前兆方法进行了更加系统、深入的研究, 主要是对云南和首都圈地区的 20 个观测点的全部观测资料进行了清理分析, 研究了地热数据处理、异常提取、三要素预报方法和预报效能评价的方法, 取得了多方面的研究成果, 从而深化了对地热方法及其机理的认识 (国家地震局预测预防司, 1997)。

我国的地热前兆台网自正式运行以来, 产生了大量纸介质和电子格式的地热前兆观测数据。按每台地热前兆仪器每年所记录的分钟值数据约 52 万个计算, 全国 200 多个地热前兆台站若干年的观测数据量是巨大的。为了充分利用这些数据, 全面、系统地推动地热前兆方法的不断发展, 需

要建立起专门用于数据分析的地热前兆应用数据库, 对所有的地热前兆数据进行时间和空间上的扫描, 从这些前兆数据中寻找前兆异常信息, 进行地热方法研究。

在这样的前提下, 有必要建立地热前兆应用数据库, 并考虑对数据库的访问速度。我们通过对几种数据库平台和接口方案的对比, 最后采用 Oracle 数据库调用接口 (Oracle Call Interface, 简称 OCI) (何雄等, 2006) 的方式, 实现了地热前兆应用数据库的建立和访问。

1 地热前兆应用数据库平台和访问接口的选择

1.1 地热前兆应用数据库平台的选择

为了高速、有效地利用地热前兆数据, 首先要选择好数据库平台。目前主流的数据库平台包括 Oracle、DB2、SQL Server 等, 这三种数据库的主要特点见表 1 (Bobrowski 2000)。

通过比较可以看出, Oracle 适用于高性能要求的场合, 而 DB2 适用于海量数据存储。由于地热前兆方法研究对性能要求较高, 而且 Oracle 可以跨平台, 所以我们选择 Oracle 数据库来存储地热前兆台网的观测数据。

* 收稿日期: 2007-11-08.

基金项目: 中国地震局 2007 年度地震行业科研专项基金 (200708045) 资助

表 1 三种主流数据库的比较

数据库名	平台依赖性	伸缩性和并行性	性能	连接方式	操作简易性
SQL Server	只能在 Windows 平台上运行	并行模型不成熟，伸缩性有限	多用户时性能不佳	只支持 Windows客户端，可以用 ADQ DAQ OLEDB ODBC连接	简单
Oracle	能在所有主流平台上运行	良好的并行性和伸缩性	性能最高	支持多种工业标准，可以用 ODBC JDBC OC等方式连接	较复杂
DB2	能在所有主流平台上运行	良好的并行性和伸缩性	性能较高	支持 ODBC JDBC等方式连接	简单

1.2 地热前兆应用数据库访问接口的选择

确定了前兆应用数据库平台后，就要选择相应的访问方式。目前主流的访问方式是 ADO (ActiveX Data Objects)。ADO是微软公司推出的基于 OLE DB (嵌入式对象链接) 的一种访问数据库的接口技术。ADO的特点主要有：支持多种数据源；使用简单；由于采用 COM技术，可以在多种语言中使用；只能在 Windows平台使用 (Scappa, 2003)。

由于访问地热前兆应用数据库的速度相当重要，我们使用一种不常用的接口方式：OCI。OCI是 Oracle提供给开发人员的最强大的接口，Oracle数据库中的每一条 SQL语句都通过 OC来执行；OC提供了对 Oracle数据库所有功能最全面的访问；OCIAP中包含了数据库的最新性能、可伸缩性和安全性特性。总体上来说，OC是 Oracle提供给开发人员的功能最强大、最底层的调用接口 (Michae等, 1998)。ADO是基于 OLE DB的，而 Oracle OLE DB Provider是用 OC实现的，由于减少了中间层的调用，OC比 ADO速度更快，可以使用数据库中的所有功能，能够跨平台 (因为用 C语言接口)，而且所消耗的系统资源较少。

1.3 OC接口的自定义封装

OC是由 Oracle提供的用于开发前端应用程序及中间件的 C接口库。由于 OC是 C语言接口，要求使用者对 C语言比较熟悉，函数调用复杂，一个函数最多有十几个参数 (谭浩强, 2005)，而且很多参数类型是 Oracle自定义的类型，不太容易理解，使用起来很不方便。

由于现在地震前兆数据都使用 Oracle数据库存储，为了实现代码重用，笔者将 OC的 C接口用 C++类封装成 QEnvironment QStatement QDirectPathLoader和 QOCException四个类，这些类实现了 OC的基本功能 (Mehnick 2005)，包括直接

路径装载。通过使用这些类大大简化了 OCI编程及调试工作，提高了编程效率。现将这几个自定义类的主要功能简述如下：

QEnvironment OC环境类：用来初始化 OC环境。它只有构造函数和析构函数。构造函数原型为：QEnvironment (string username, string password, string password), username、password和 db分别对应地热前兆应用数据库的用户名、密码和地热前兆数据库实例名。String是标准 C++模板类，考虑到跨平台应用，没有使用 MFC (微软基础类) 的 CString类。

QStatement OC陈述类：用来执行 SQL语句。它的主要功能包括：初始化 SQL语句，绑定输入变量，定义输出变量，执行 SQL语句，获取 SQL语句返回的结果集。

QDirectPathLoader 用于直接路径装载，可以用它实现一张地热前兆仪器数据表的直接路径装载。

QOCException OC产生错误时抛出的异常类，比如连接前兆数据库失败或装载的前兆数据不合法时会抛出该异常。

2 地热前兆应用数据库的建立和访问

2.1 地热前兆数据的收集与整理

地热前兆观测数据来源有：① 是中国地震台网中心的地震前兆数据库；② 是 2000年以前由观测台站上报的纸介质数据；③ 是从各省局或台站收集的前兆数据 (中国地震局监测预报司, 2003)。由于中国地震台网中心的地热前兆数据都是各省局上报的压缩包，首先要对数据包进行解压。因为包的数量巨大 (6万多个)，手工解压是不现实的，所以编写了解压工具，使用 VC调用 unrar库对压缩包解压。该程序先扫描目录，并得到目录下所有文件名，对于每一个文件，如果是

压缩文件则解压到一个临时目录，接着扫描这个临时目录的所有文件，由文件名得到测项代码：如果是地热前兆数据，就将该文件的数据内容添加到要装载入前兆应用数据库的数据文件上，这个文件的格式为“时间 数据值”，例如“199901010000 23.7982”。对于2000年前的纸介质地热数据，聘请兼职录入人员进行录入。为了保证录入的准确和高效，使用VC和Matlab（张志涌，2003）编写了录入程序，对输入的前兆数据做均方差检验，并可以绘制任意天的曲线供录入人员校对。

2.2 前兆应用数据库表的建立

地热前兆应用数据库初始化后要建立4张表：台站表、仪器信息表、日志表和数据区间表。台站表用来保存所有的台站信息，包括台站名、台站代码、台站经纬度等信息；仪器信息表用来保存所有的仪器信息，主要包括仪器ID、仪器类型、所属台站代码等信息；日志表用来保存所有的台站日志信息；数据区间表用来统计所有仪器数据的续存时间段，包括表名、数据开始时间、数据结束时间。

为了提高地热前兆数据表的查询和访问速度，将所有的地热前兆数据按仪器分类，每个仪器一张数据表，表名约定为“T台站代码_手段代码”，前兆数据表在程序中动态创建。对于地热前兆数据存储有两种方式：一行存放一天的数据或者一行只存放一个数据。采用前一种方式可以节省约一半的空间，但由于一天的数据在一行上，数据操作的对象是以天为单位，必须要把单个数据构造成一天的数据流（对应于更新）或者把一天的数据流解析为单个数据（对应于查询），这必然影响数据处理的速度。后一种方式不需要构建或解析的转换，而且能有分钟粒度的控制，所以使用了后一种方式保存地热前兆数据，这也是一种以空间换效率的做法。

2.3 地热前兆数据装载

地热前兆观测数据的准备工作完成后，就可

以将这些数据装载到前兆应用数据库中。Oracle提供了一种直接路径装载方式，可直接写表的数据文件，这是最快的数据加载方式，但由于它绕过了触发器和约束条件等，装载之前必须保证数据是正确的（即不违反表中的所有约束条件）。使用直接路径装载时，客户端首先分配数据缓冲区，将数据存放到数据缓冲区，然后分配列数组，将每行数据设置在对应的列数组行上，装载时OC把列数组转换成数据流传送到服务器端，服务器端将数据流重新转换成列数组，再按照指定的格式将列数组加载到表中。具体流程如图1所示。

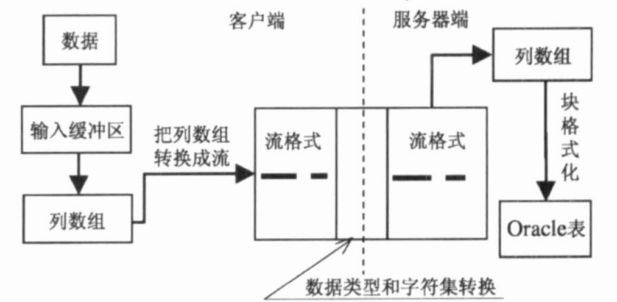


图 1 直接路径装载示意图

直接路径装载可以在程序中调用Oracle的sqlldr装载工具，但这种方法缺少对程序的控制，比如某个数据文件内容不规范导致装载失败时无法知道这个错误的细节（例如是文件的哪一行导致错误），所以应用OC的直接路径接口进行装载（Kruglinsk等，1999）。

在程序中用自己封装的QDircPathLoader类实现装载，主要步骤如下：① 初始化QEnvironment对象；② 初始化QDircPathLoader对象；③ 设置列数组中的数据；④ 将列数组转换为数据流；⑤ 装载数据流；⑥ 如有必要，重复步骤③~⑤；⑦ 结束装载。图2是用直接路径装载1台仪器数据文件的流程图。表2是4个台站地热前兆数据文件分别用OC和ADO装载消耗的时间对比。

表 2 ADO和OC装载用时对比

台站名	文件行数	ADO批量装载用时 / s	OC直接路径装载用时 / s
四川攀枝花仁和	1 069 920	2 412	30
四川石棉	1 025 280	2 205	29
四川西昌太和	1 064 160	2 362	31
云南下关 1 台	1 023 840	2 160	30

注：测试时的服务器环境为双英特尔奔腾 III 650MHz 处理器，512M 内存，Oracle9 数据库。
?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

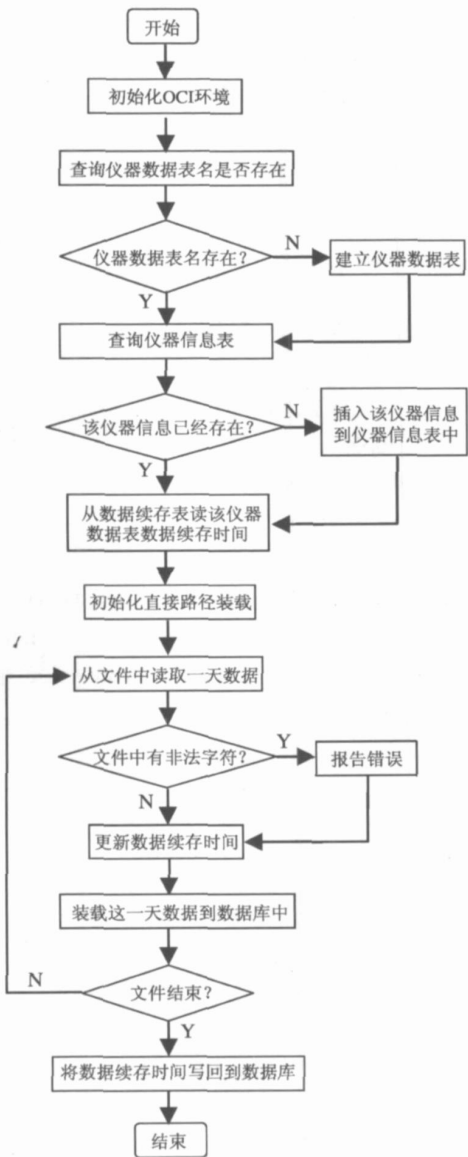


图 2 用直接路径装载前兆数据文件的流程图

2.4 地热前兆应用数据库的访问

为了提高查询效率，使用 OC 的数组接口实现，使用 SQL 语句执行数据的查询、更新、插入和删除操作。OC 的数组接口能在与服务器的交互中完成操作，而不是反复执行某条 SQL 语句。OC 能够最大程度地控制程序的运行，执行 Oracle 服务器所允许的所有数据库操作，访问 Oracle 数据库服务器里的所有数据类型，以引用的方式访问对象及其元数据，动态获取、修改对象的属性值。

OC 数据访问包括以下几个部分：初始化 OCI 环境；分配所需的句柄与数据结构，建立与服务

器的连接和用户会话；准备 SQL 语句；通过在数据库服务器上执行 SQL 语句与数据库服务器进行数据交换，并进行必要的数据处理；重复执行准备就绪的语句或准备一个新的要执行的语句；中止会话，服务器断开；释放句柄和数据结构。

OC 将 SQL 语句的参数分为“绑定”和“定义”两种。绑定对应输入变量，“定义”对应输出变量。

下面是从数据库读取地热数据的主要步骤：
① 初始化 QEnvironment 对象；
② 初始化 QStatement 对象；
③ 初始化 SQL 语句，申请接收缓冲区；
④ 定义查询结果要保存到的缓冲区；
⑤ 执行查询。表 3 是分别用 ADO 和 OC 访问 4 个地热台站观测数据所用时间对比。

表 3 ADO 和 OC 查询用时对比			
台站名	表行数	ADO 查询用时 / s	OC 查询用时 / s
四川攀枝花仁和	1 069 920	12	2
四川石棉	1 025 280	13	3
四川西昌太和	1 064 160	14	4
云南大关 1 台	1 023 840	12	3

注：测试时的服务器环境为双英特尔奔腾 III 650MHz 处理器，512M 内存，Oracle9i 数据库。

3 结语

从表 2 和表 3 可以看出，使用 OC 能大幅度提升表的装载速度和查询速度。所以，在只使用 Oracle 数据库、对性能要求很高或者要使用数据库的所有功能时，使用 OC 是最合适的。此外，OC 对 Oracle 数据库的性能调整也非常重要。

参考文献：

付子忠. 1990 云南省地热前兆台网及地热前兆方法技术的研究：主要研究实验报告 [R]. 北京：国家地震局地壳应力研究所.
国家地震局预测预防司. 1997. 地下流体地震预报方法 [M]. 北京：地震出版社.
何雄. 2006. Oracle Spatial 与 OCI 高级编程 [M]. 北京：中国铁道出版社.
谭浩强. 2005. C 程序设计（第 2 版）[M]. 北京：清华大学出版社.
张志涌. 2003. 精通 MATLAB6. 5 版 [M]. 北京：北京航空航天大学出版社.
中国地震局监测预报司. 2003. 地震前兆数字观测公用技术与台网 [M]. 北京：地震出版社.
Bobrowski. 2000. Oracle 8 体系结构 [M]. 王琰，王磊，蒋蕊译. 北京：机械工业出版社.

Kruglinski D J Wingo S Shepherd G 1999. Visual C++ 6.0技术内幕 [M]. 朱继满, 蒋方帅, 懂柱, 等译. 北京: 希望电子出版社.

Mehnick J 2005. Oracle Call Interface Programmers Guide 10g Release 2 (10.2) [CP/DK].

Michael A Michael J C 1998. Oracle8初学者指南 [M]. 钟鸣, 刘晓霞, 译. 北京: 机械工业出版社.

Seppa D 2003. ADO.NET技术内幕 [M]. 梁超, 张莉, 贺堃, 译. 北京: 清华大学出版社.

Performing High-speed Access to Geothermal Precursor Database by Wrapping Oracle Call Interface in C++ Class

WANG Jun¹, ZHAO Gang¹, HE An-hua¹, GUO Miao-xi¹, GUO Bo-lin¹, QIN Jiu-gang²

(1. Institute of Crustal Dynamics, CEA, Beijing 100085, China)

(2. National Earthquake Infrastructure Service, CEA, Beijing 100036, China)

Abstract

Geothermal precursor method is a relatively new earthquake prediction method. It is very necessary to collect all available geothermal data and construct geothermal database for a systematic research on this method. We presented a way it can load and access geothermal data at high speed by wrapping oracle call interface in four C++ classes, and also improve the readability and maintenance of source code.

Key words: geothermal precursor; database; OCI; Oracle; C++; direct path load



张 昱 甘肃省地震局高级工程师。1986年毕业于西安地质学院,获工学学士学位。主要从事地下流体地震分析预报等方面的研究。



何案华 中国地震局地壳应力研究所助理研究员。2005年毕业于中国地震局地壳应力研究所,获地球物理硕士学位。主要从事地震前兆仪器的研制和网络技术工作。



史红军 长春市地震局榆树地震监测台工程师。1996年毕业于中国地质大学(北京),主要从事地震监测工作。



王 博 中国地震局地壳应力研究所在读硕士研究生。2005年毕业于聊城大学环境与规划学院,获理学学士学位。现主要从事地震流体动力学与地震分析预报研究工作。



王 军 中国地震局地壳应力研究所在读硕士研究生。主要从事地震前兆数据分析和软件设计工作。



张丽芬 中国地震局武汉地震研究所研究实习员。2003年华东地质学院资源勘查工程专业,获学士学位。2006年毕业于东华理工大学地球化学专业,获硕士学位。参与并承担了多项水库地震相关研究项目,包括国家科技支撑项目、国家自然科学基金项目、地震研究所科技发展基金项目等。主要从事水库诱发地震工作。



夏静瑜 云南大学地球物理系在读硕士,主要从事地球物理学方面的研究工作。