

大型数据平台搭建技术在地震行业中的应用与研究*

单德华, 杨红艳, 孙鸿雁, 吴野

(辽宁省地震局, 沈阳 110031)

摘要: 依据地震行业业务系统的数据处理和存储需求, 提出了数据模型和数据平台建设思路。结合地震行业网中数据平台的建设, 在国家中心、区域中心和台站节点分别采用了相应的数据库平台和数据存储平台技术, 设计了各类节点的数据备份策略, 实现了满足海量数据流高效运行的大型数据平台。

关键词: 数据库; 存储结构; 数据备份

中图分类号: P315.69

文献标识码: A

文章编号: 1000-0666(2010)02-0234-04

1 地震行业数据平台建设需求分析

地震行业业务系统数字化和网络化建设完成后, 在日常工作中产出了大量的地震专业数据, 这些数据的存储和应用需求越来越复杂, 其需求的满足依赖于在地震行业网中搭建能够快速处理海量数据的数据平台(中国地震局, 2005a)。

地震行业业务系统以及相关数据分类主要有:

(1) 测震类数据: 数字地震台网产出的连续波形、事件波形、地震震相、震源机制解等观测数据和处理结果, 以及台站参数、仪器参数等基础数据(中国地震局, 2005b)。

(2) 强震类数据: 数字强震台网实时观测的有震记录; 国内外有影响地震的强震观测数据和处理结果。

(3) 前兆观测类数据: 地震前兆台网产出的观测数据、产品数据, 以及台站参数、仪器参数等基础数据(中国地震局, 2005c)。

(4) 分析会商类数据: 各省市地震局、各地震台站的分析会商相关信息。

(5) 震情和灾情类数据: 国内外震情和灾情信息。

(6) 地震目录类数据: 根据数字地震台网的记录编制的国内外大震速报目录, 全国月报目录、快报目录、临时地震序列目录, 全球地震目录等。

(7) 运行管理类数据: 各级节点汇总的系统运行数据, 与运行相关的人员、场地、设备情况数据、观测点数据和数字资产数据等。

(8) 综合地理信息类数据: 基础地理信息数据。

(9) 政务信息类数据: 地震行业内部各级地震部门上传至中国地震局和中国地震局下传的行政、公共信息以及网上办公事务的数据信息。

(10) 宣传教育类数据: 汇集到本地信息节点的防震减灾常识和法规, 地震科普知识宣教多媒体信息。

(11) 用户属性和资源权限: 整个“观测网络”软件系统中用户的信息编码规范以及相关的部门、机构、组织和系统的编码规范, 用户信息与数据库数据资源的访问权限的对应关系, 用户数据的格式, 用户信息交换的相关规范等。

(12) 监控信息: 台站仪器和网络设备的日志、状态、控制参数等信息。

(13) 元数据: 地震观测网络各个业务和学科数据的元数据信息。

2 地震监测系统数据模型建立

数据建模主要解决信息系统的存储结构、效率、方式等问题, 并进行数据结构设计, 将模型进行推广和部署。数据建模的主要任务包括数据

* 收稿日期: 2009-04-08.

基金项目: 地震信息网络安全运行、风险防范与重大事件处置技术研究(地震行业科研专项 200708015)资助.

存储、数据建模和物理建模（刘娜等，2009；梁丽芳，白海丽，2005）。

数据存储：主要包括数据存储分区、数据访问接口、数据库用户/角色设计、数据库优化、数据备份、数据迁移、数据灾难恢复。其他数据永久存储手段包括光盘库、带库、磁盘阵列等。

数据建模：对信息、测震和前兆学科的基础数据和观测数据进行统一的数据建模，主要是 E-R 模型的分析 and 建模（萨师焯，王珊，2000；刘政凯等，2007）。

物理建模：对关系型数据库而言，包括表、存储过程、触发器、各种数据管理的脚本、元数据管理（ShaShi Shekhar, Sanjay Chaqla, 2004；Ken Henderson, 2002）。

地震监测系统数据建模按业务系统可分为测震、前兆、信息分项数据建模。在数据建模过程的各个阶段，为各分项提供数据建模的方法、规范的意见和建议，规划、汇总各阶段产出文档，完成公共信息和建模要素的建设任务，并在各分项数据模型建设的基础上，完成各级节点的数据库的物理部署，实现脚本、原形测试与试点工作，维护地震监测系统数据库的正常运行。按照数据模型设计的方法和定义，通过数据模型设计的需求调研、逻辑设计、物理设计、数据模型的试点和测试，共同完成数据建模的建设任务。数据建模流程见图 1。

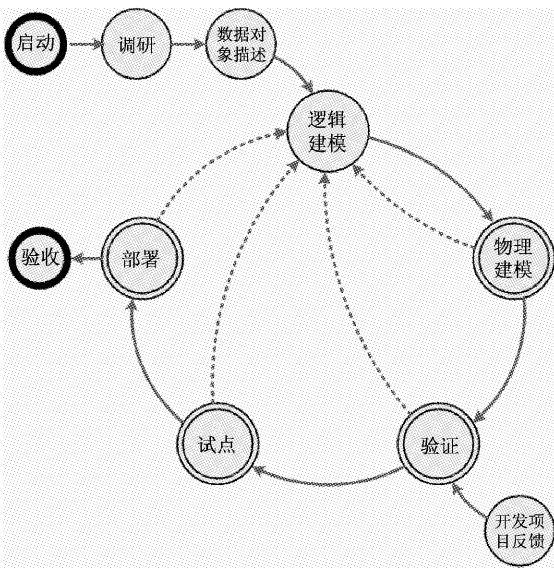


图 1 数据建模流程图

Fig. 1 Data modeling flow chart

3 地震行业数据平台设计与实现

国家中心通过 150TB 的 SAN + NAS 存储系统和备份系统搭建台网中心数据系统的基础设施，完成汇集、入库等数据库应用系统的开发和部署，建成“十五”观测数据集成存储和管理的平台，系统每天汇集、存储、管理的观测数据量达到 35 GB 以上，对 1 021 个测震台站和 70 多个前兆台站的数据进行汇集和管理，数据库容量超过 1 500 万条。

区域中心通过 8.5TB 的 SAN + NAS 存储系统搭建区域中心数据系统的基础设施，完成汇集、入库等数据库应用系统开发和部署，建成“十五”区域观测数据集成存储和管理的平台，每个区域中心平均每天汇集、存储、管理的观测数据量达到 2 GB 以上，对区域内测震台站和前兆台站数据进行汇集和管理，数据库容量超过 30 万条。

大中城市节点、县级节点和台站节点采用 144 GB 的 PC 服务器搭建本节点范围内的地震观测仪器数据采集和交换平台。有些地方还将本地区及临近地区的测震实时数据汇集到节点的服务器中，形成虚拟观测台网。每个节点平均每天数据流量超过 400 MB。

通过在国家中心和区域中心安装双机 Oracle 数据库，实现应用集成、综合共享和数据建模功能；在台站安装单机 Oracle 数据库，用于存放综合在线监控类数据、前兆观测数据及其元数据，也可根据需求选择存放分析预报工作所需的其它基础数据（如地震目录）和产出结果、速报台所需的测震观测数据和元数据等；在大中城市节点安装单机 Oracle 数据库，用于存放本市及临近地区的震情和灾情数据、地震信息服务系统汇集的综合信息、应急指挥系统汇集的数据等；在县节点安装单机版 Oracle 数据库，用于存放本地及临近地区的震情和灾情数据、地震信息服务软件必要的后台基础数据等。在国家中心和区域中心安装的数据库采用 Oracle 10 G 的 RAC 技术。

3.1 国家中心数据库系统

中国地震台网中心信息服务系统由应用集成数据库、综合数据库、双机数据库 3 个物理数据库系统组成，实现专业数据的管理。数据库软件

采用企业版 Oracle 10 G，硬件采用 IBM 小型机（图 2）。

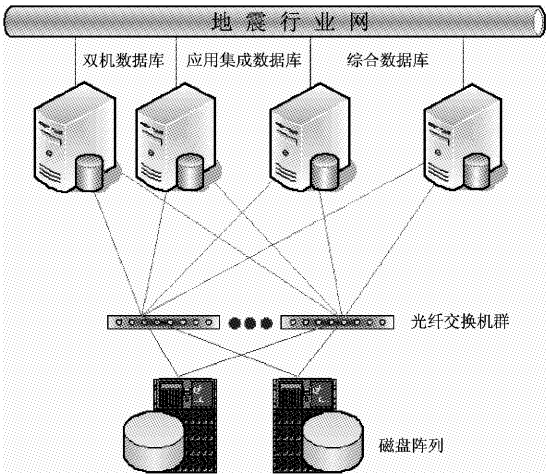


图 2 中国地震台网中心信息服务系统
Fig. 2 National center information service system

3.2 区域中心数据库系统

区域中心采用双机 RAC 方式构建数据库系统。业务库和综合库放在一起，区域中心数据库软件采用企业版 Oracle 10 G，硬件平台采用 SUN、IBM、曙光三类小型机（图 3）。

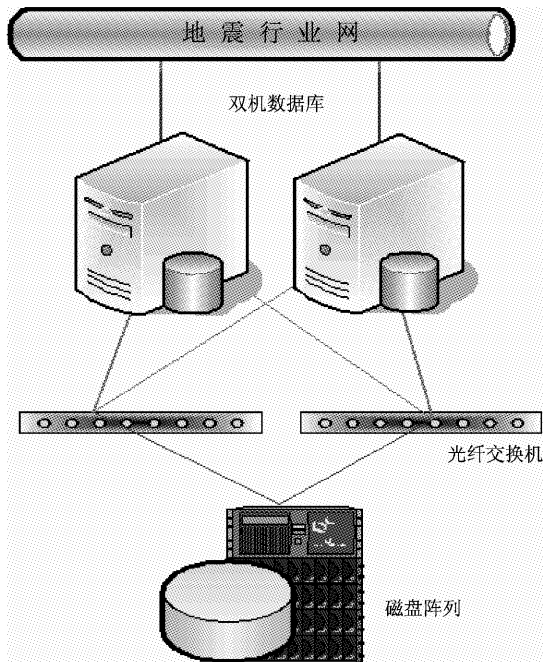


图 3 区域中心信息服务系统数据库系统
Fig. 3 Region center information service database system

3.3 大中城市、县级和台站节点数据库系统

台站、大中城市和县级节点采用单机安装标准版 Oracle 数据库系统，硬件采用机架式 PC Server（图 4）。

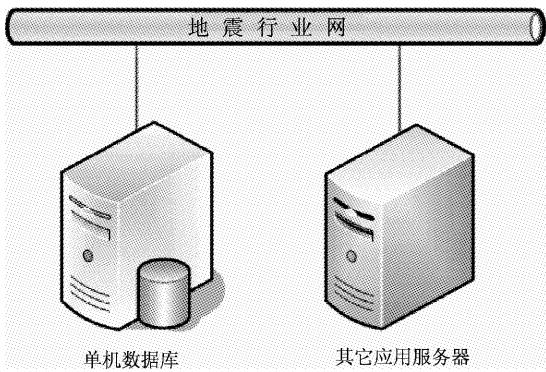


图 4 台站、大中城市和县级节点数据系统
Fig. 4 Stations, cities and county-level node data system

台站、大中城市和县级节点数据库安装在单机服务器上，服务器操作系统为 Suse Linux，数据库系统为 Oracle，每个节点有 1 台数据库服务器和多台应用服务器。

4 建设成果及特色

4.1 国家—区域—台站三级分布式数据库系统

在 1 个国家中心和 41 个区域中心采用 Oracle RAC 双机系统，具有高可靠性、负载均衡和故障自动切换机制同时也为海量数据的准实时接收和汇聚的数据提供了可靠的数据容器。大中城市、县级节点、台站节点均采用标准版 Oracle。

在国家中心和区域中心，基于测试和验证，设计了测震学科的 10 分钟等时的数据结构（原始数据格式为 512 Byte），使数据纪录数缩小为原来的 1/200，可提供 3 个月在线波形数据的高效存储和有效访问；采用了基于日期映射的分区技术，BLOB 和 CLOB 存储类型单独存储的策略提高了数据汇集入库、存储管理和检索数据访问效率。

4.2 为海量数据处理奠定基础

全行业网运行海量数据采集、传输、存储、共享和服务模式。“十五”期间建成的各种观测台网产生的原始观测数据量为每年 20 ~ 30 TB，针对这种情况，设计了数据存储管理生命周期，按照

服务层、调度层和存储层的 3 个数据服务层次搭建服务应用系统原型，实现了海量数据的长期在线服务。

4.3 实现观测数据多级数据备份

全网实现了磁盘级、系统级、数据库级、文件级、归档级等多级备份功能，可有效解决海量地震观测数据的备份和恢复。根据数据存储、数据库使用的特点，针对不同的应用模式和数据库系统执行不同的备份策略。

参考文献：

梁丽芳,白海丽,石金峰. 2005. 利用 CASE 工具建立地理数据库的方法[J]. 辽宁工程技术大学学报,24(增刊):56-58.
刘娜,谢英情,马颖鹏,等. 2009. 基于 GeoDatabase 的地震构造背景

场数据库设计与实现[J]. 地震研究,32(1):94-98.
刘政凯,王伟,张荣,等. 2001. 地震图纸数字化及数据库管理系统[J]. 地震学报,(3):31-34.
萨师焯,王珊. 2000. 数据库系统概论[M]. 北京:高等教育出版社.
郑文锋,银正彤. 2007. 基于 OpenGIS 的地震信息采集与震灾评估系统[J]. 地震研究,30(2):196-201.
中国地震局. 2005a. 中国地震信息服务系统技术规程[M]. 北京:地震出版社.
中国地震局. 2005b. 中国数字测震台网技术规程[M]. 北京:地震出版社.
中国地震局. 2005c. 中国数字前兆台网技术规程[M]. 北京:地震出版社.
Ken Henderson. 2002. SQL Server 存储过程、XML 和 HTML 高级指南[M]. 北京:清华大学出版社.
ShaShi Shekhar, Sanjay Chakla. 2004. 空间数据库[M]. 谢昆青, 马修军, 杨冬青, 译. 北京:机械工业出版社.

Application and Research for Large-scale Data Platform
Build Technology in Earthquake Profession

SHAN De-hua, YANG Hong-yan, SUN Hong-yan, WU Ye
(Earthquake Administration of Liaoning Province, shenyang 110031, Liaoning, China)

Abstract

Based on industry seismic data processing and storage demands, this paper presents a data model and data platform construction ideas. For seismic data in the industry network construction platform at the national centers, regional centers and stations, a corresponding node of the database platform and data storage platform technology is adopted. The design of a variety of node data backup strategy, implementation of the massive data stream is adopted to satisfy the large-scale high-performance data platform.

Key words: database, memory structure, data backup