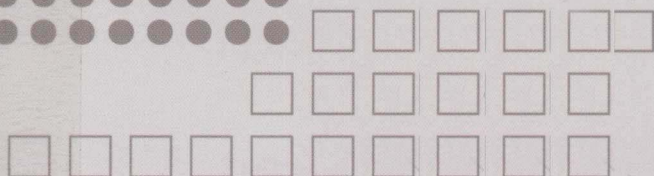
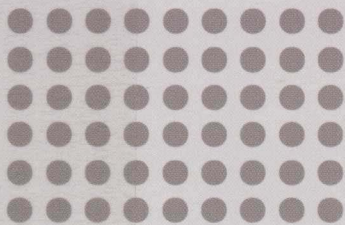
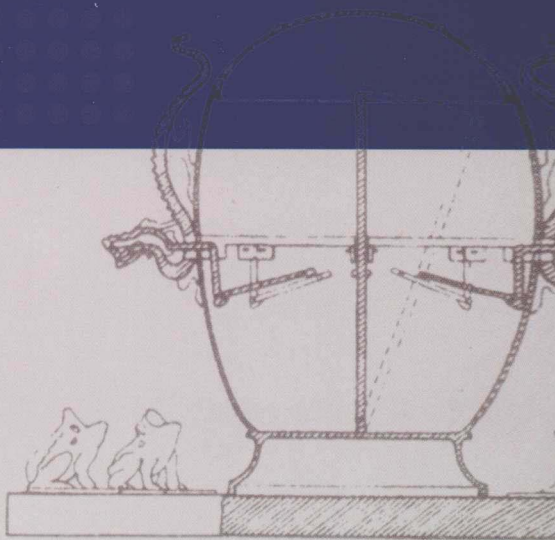


数据建模




王方建 主 编



数据建模

王方建 主 编

 科学出版社

内 容 简 介

本书以中国地震监测系统为应用案例,由浅入深介绍数据建模方法及其相关技术。包括数据模型命名规则、设计规范、建模技术、工作组织与流程规范等,并介绍了这个项目的数据库备份与恢复方案、数据中心的部署规划等技术,让读者不仅可以集中地学习数据建模的主要理论方法,而且可以了解实际应用中解决问题的操作方法。

本书可作为大型网络系统数据库建模的教学、实践和应用的参考书,也可作为相关工作人员的学习资料。

图书在版编目(CIP)数据

数据建模 / 王方建主编. — 北京: 科学出版社,
2012. 2

ISBN 978-7-03-033344-5

I. ①数… II. ①王… III. ①数据模型—建立模型
IV. ①TP311.13

中国版本图书馆CIP数据核字(2012)第005990号

责任编辑: 赵东升 于先军 / 责任校对: 杨慧芳
责任印刷: 新世纪书局 / 封面设计: 丁世杰

科学出版社 出版

北京东黄城根北街16号

邮政编码: 100717

<http://www.sciencep.com>

中国科学出版集团新世纪书局策划

北京市鑫山源印刷有限公司印刷

中国科学出版集团新世纪书局发行 各地新华书店经销

*

2012年1月第一版
2012年1月第一次印刷
字数: 243 000

开本: 16开
印张: 10

定价: 38.00元

(如有印装质量问题, 我社负责调换)

前 言

中国数字地震监测系统在各专业软件开发和数据管理层面必须对监测系统的数据存储、管理和应用体系进行数据设计，进行数据设计主要目的就是提供完备和规范的数据模型，为关键业务数据（包括各种观测数据和数据产品等）的存储和各个业务软件的开发实施提供可靠的数据模型，实现数据模型的统一规划和一致性，保障系统的建设和后续的改进工作。

数据设计工作的基础性和支撑作用，决定了其任务必须优先进行，尤其是基础数据和业务间公用的数据。因此，本书的创作初衷是尽可能清楚地为读者解释如何分析、设计及实现地震监测业务系统的数据库。实际上本书对进行理论研究的人士也非常有用，而且简单明了地阐述了数据库设计的过程。

本书共分 13 章，主要内容如下：

第 1 章到第 6 章介绍地震监测业务系统的数据库设计的规范和方法，包括数据设计指南、命名规则、设计规范、物理数据模型设计规范、数据建模技术与方法和工作组织与协作流程规范。

第 7 章介绍地震监测业务系统数据库访问、调优和资源管理策略。

第 8 章介绍数据平台框架设计方案，包括数据系统现状分析、数据平台框架和业务应用的整合模式选择。

第 9 章介绍数据库备份与恢复方案。

第 10 章介绍地震监测系统数据平台的物理模型设计，包括设计概述、物理模型设计说明和原型系统设计。

第 11 章介绍数据逻辑模型设计方案。

第 12 章介绍国家中心数据平台部署规划，包括数据库部署架构、数据库同步与数据交换策略和硬件配置。

第 13 章介绍区域中心数据平台部署规划。

本书内容上重点突出，强调实用，概念上简洁准确，清楚易懂，适合地震相关工作人员的学习。本书编写过程中，还参考了国内外专家和学者的著作和教材，在此一并表示感谢。面对技术发展的迅猛势头，深感自己才疏学浅，对于书中存在的错误和不妥之处恳请专家学者和广大读者不吝指正，并真心希望和广大读者互动过程中能得到提高。

目 录

第 1 章 数据设计指南.....	1
1.1 数据设计原则.....	2
1.2 数据设计工具.....	2
1.2.1 工具选择策略.....	2
1.2.2 可以使用的设计工具.....	3
1.2.3 推荐使用的设计工具.....	3
1.3 数据设计指南.....	3
1.3.1 概念数据模型 (CDM) 的建模.....	3
1.3.2 物理数据模型 (PDM) 的建模.....	4
1.4 相关数据设计规范约束.....	4
第 2 章 数据模型命名规则.....	6
2.1 建立命名规范的目的.....	7
2.2 命名规范说明.....	7
2.2.1 总体原则.....	7
2.2.2 数据模型版本命名规范.....	7
2.2.3 逻辑模型命名规则.....	8
2.2.4 物理模型命名规则.....	8
第 3 章 数据模型设计规范.....	10
3.1 数据模型开发方法和工作流程.....	11
3.1.1 概念.....	11
3.1.2 开发方法.....	11
3.1.3 工作流程.....	12
3.2 词汇表开发规范.....	13
3.2.1 方法.....	13
3.2.2 管理.....	14
3.3.3 规范.....	14
3.3 实体对象模型开发规范.....	14
3.3.1 方法.....	14
3.3.2 管理.....	15
3.3.3 规范.....	15
3.4 逻辑数据模型开发规范.....	16
3.4.1 方法.....	16
3.4.2 管理.....	17

3.4.3	规范.....	17
第4章	物理数据模型设计规范.....	18
4.1	物理数据模型概念.....	19
4.1.1	设计约束.....	19
4.1.2	基本概念.....	19
4.2	物理数据模型开发方法.....	19
4.2.1	规划 E-R 模型.....	20
4.2.2	用户数据评估.....	23
4.2.3	规划物理数据库.....	23
4.2.4	生成 DDL.....	27
4.2.5	生成数据字典.....	27
4.2.6	优化物理数据模型.....	27
4.2.7	海量、高增长数据存储设计的建议.....	27
4.3	物理数据模型管理.....	28
4.4	物理数据模型命名规范.....	29
4.5	物理数据模型规范.....	29
4.6	表空间设计原则.....	29
4.6.1	表空间划分原则.....	29
4.6.2	表空间物理存储分布考虑.....	30
4.6.3	表空间相关的参数.....	30
第5章	数据建模技术与方法.....	32
5.1	工作流程.....	33
5.1.1	需求分析.....	33
5.1.2	数据对象描述.....	33
5.1.3	逻辑数据模型.....	33
5.1.4	物理模型.....	33
5.1.5	验证和试点.....	34
5.1.6	部署推广.....	34
5.2	数据建模技术流程.....	34
5.2.1	数据建模迭代过程.....	34
5.2.2	数据模型工具选择.....	35
5.2.3	逻辑建模.....	35
5.2.4	物理建模.....	36
5.3	数据库设计.....	36
5.3.1	部署设计.....	36
5.3.2	数据库运行管理和优化设计.....	36
5.3.3	存储和备份设计.....	36

5.3.4	数据服务接口	36
5.4	验证与试点	36
5.5	协作流程	37
第 6 章	工作组织与协作流程规范	38
6.1	数据模型建设组织流程	39
6.2	协作流程	40
6.2.1	首次提交流程	40
6.2.2	迭代流程	41
6.2.3	变更流程	41
第 7 章	地震监测业务系统数据库管理策略	44
7.1	访问策略	45
7.1.1	用户管理策略	45
7.1.2	程序访问策略	45
7.2	调优策略	46
7.2.1	结构调优	47
7.2.2	SQL 调优	49
7.3	资源管理策略	49
7.3.1	规划原则	49
7.3.2	规范说明	50
第 8 章	数据平台框架设计方案	51
8.1	数据系统现状分析	52
8.1.1	数据分布	52
8.1.2	数据集	53
8.1.3	数据流	53
8.2	数据平台框架	57
8.2.1	数据平台架构设计	57
8.2.2	数据整合模式	59
8.3	业务应用的整合模式选择	61
8.3.1	测震	61
8.3.2	前兆	64
8.3.3	信息	64
第 9 章	数据库备份与恢复方案	67
9.1	数据库备份与恢复基础知识	68
9.1.1	逻辑备份	68

9.1.2	物理备份.....	68
9.2	数据库备份与恢复方案.....	69
第 10 章	地震监测系统数据平台的物理模型设计.....	72
10.1	设计概述.....	73
10.1.1	设计范围.....	73
10.1.2	整体框架.....	73
10.2	物理模型设计说明.....	75
10.2.1	测震数据缓存.....	75
10.2.2	前兆数据缓存.....	82
10.2.3	机构、用户.....	85
10.2.4	台站仪器参数.....	92
10.2.5	元数据.....	116
10.2.6	归档数据/文件.....	116
10.3	原型系统设计.....	117
10.3.1	数据量分析.....	117
10.3.2	表空间设计.....	119
第 11 章	数据逻辑模型设计方案.....	120
11.1	逻辑模型.....	121
11.1.1	实体列表.....	121
11.1.2	实体关系描述 (E-R 图).....	122
11.2	数据对象描述.....	123
11.2.1	公共信息数据对象描述表.....	123
11.2.2	仪器信息数据对象描述表.....	126
11.2.3	测震信息数据对象描述表.....	130
11.2.4	前兆信息数据对象描述表.....	135
第 12 章	国家中心数据平台部署规划.....	138
12.1	数据库部署架构.....	139
12.2	数据库同步与数据交换策略.....	140
12.2.1	测震分项与信息分项数据同步.....	140
12.2.2	前兆分项与信息分项数据同步.....	141
12.2.3	信息分项中数据存储.....	142
12.3	硬件配置.....	142
第 13 章	区域中心数据平台部署规划.....	143
13.1	数据库部署架构.....	144
13.2	数据库同步与数据交换策略.....	144

13.2.1 测震分项与信息分项数据同步	144
13.2.2 前兆分项与信息分项数据同步	147
13.2.3 信息分项中数据存储	147
13.3 硬件配置	147
参考文献	149

第 1 章

数据设计指南

中国数字地震监测系统在各专业软件开发和数据管理层面必须对监测系统的数据存储、管理和应用体系进行数据设计，进行数据设计的主要目的就是提供完备和规范的数据模型，为关键业务数据（包括各种观测数据和数据产品等）的存储和各个业务软件的开发实施提供可靠的数据模型，实现数据模型的统一规划和一致性，保障系统的建设和后续的改进工作。

数据设计工作的基础性和支撑作用，决定了其任务必须优先进行，尤其是基础数据和业务间公用的数据。

1.1 数据设计原则

1. 公用数据进行统一设计和规划

对系统内的基础数据和业务公用的数据进行统一的建模工作，减少重复劳动，保证整个系统数据模型和物理数据模型的一致性。

2. 模型转化

主流的设计工具都支持从数据模型向物理数据模型的转换和反向的转换。

3. 正向工程

新设计的业务软件，坚持先进行数据模型的建模，使用工具转换为基本的物理数据模型，根据需要修改和完善物理数据模型。

4. 反向工程

对于已有的物理数据模型或物理数据库中的永久存储元素，或者是数据库 SQL 脚本（如 DDL），如果没有进行数据模型的建模工作，可以使用反向工程，转换为数据模型，根据需要修改和完善数据模型。

1.2 数据设计工具

1.2.1 工具选择策略

1. 纸上作业

理论上如果对 OOAD 和 RDBM 建模的各种方法和技术都比较熟悉，可以不使用任何 CASE 工具进行数据设计，手工进行也可以，不过工作效率是一个问题。

2. 自行选择合适工具

设计人员可以选择自己最熟悉的设计工具，但必须保证最终的工作成果可以整合，否则，必须使用统一的设计工具。

3. 使用通用 CASE 工具

与整个开发过程结合，使用统一的设计和建模工具，方便协同工作和配置管理工作。提高工作效率，降低学习和培训成本。

4. 使用专用 CASE 工具

可以统一采用专用的数据设计工具或 ORM 工具进行设计工具，主要优点工具本身功

能比较强大。

5. 使用已有 CASE 工具

如果行业内已经有数据设计和建模工具，为降低采购成本和培训成本，可以沿用已有的建模工具。

1.2.2 可以使用的设计工具

1. UML 建模工具

如 IBM Rational 的有关产品 (Rose)，Borland 公司的 Together 等等。

2. 专用的数据建模工具

如 ERWin, PowerDesigner 的各种版本等。

1.2.3 推荐使用的工具

1. 通用建模工具推荐

考虑到各个项目的成果引用和连续的工程实施，建议采用可以提供物理模型转换统一的 UML 建模工具，如 IBM Rational Rose 的新版本。

2. 专用数据建模工具推荐

建议使用 PowerDesigner 进行数据建模和物理数据模型建模工作。

3. 优化组合方案推荐

如果资源允许，最好使用 Rose+PowerDesigner 组合工具进行数据设计。后面的规范都是针对这种方案进行描述。

1.3 数据设计指南

1.3.1 概念数据模型 (CDM) 的建模

CDM 是 PowerDesigner 工具中的 E-R 模型，CDM 的建模专指使用 PowerDesigner 工具建模而言，使用其他的建模工具，相当于 ERM (实体关系模型)

1. 正向工程建模

(1) 输入

通过业务分析过程得到的业务实体类（Business Entity）抽取。

通过需求分析过程得到的实体类（Entity）抽取。

进一步通过分析过程得到的演化后的实体类（Entity）抽取。

类设计过程中的需要永久存储的类的抽取（Class）。

(2) 输出

完成 CDM 模型的建立。

(3) 迭代

对上面的输入和输出进行多次迭代，直到满足需要。

2. 反向工程建模

(1) 输入

使用工具对 PDM 进行转换生成 CDM。

使用工具对物理数据库中的永久存储元素进行抽取转换直接生成 PDM 和 CDM。

使用工具对数据库 SQL 脚本（如 DDL）进行抽取转换直接生成 PDM 和 CDM。

(2) 输出

直接得到的 CDM。

根据设计需要修改和完善 CDM。

(3) 迭代

对上面的输入和输出进行多次迭代，直到满足需要。

1.3.2 物理数据模型（PDM）的建模

1. 输入

正向工程建模：使用工具进行正向工程转换 CDM 为 PDM，或者使用工具直接进行物理数据模型的建模工作。

反向工程建模：从物理数据库中使用工具生成物理数据模型（PDM）修改和完善物理数据模型。

2. 输出

完整的物理数据模型 PDM。

生成数据库 SQL 脚本（如 DDL）。

3. 迭代

对上面的输入和输出进行多次迭代，直到满足需要。

1.4 相关数据设计规范约束

包括以下几个数据设计方面的规范可以使用。

- 数据模型命名规范：有关数据模型的命名的原则和规范。
- 数据模型设计规范：数据模型设计方面的要求和原则。
- 物理数据模型设计规范：物理数据模型设计方面的要求和原则。
- 物理数据模型设计规格书规范：体现物理数据模型设计成果的物理数据模型设计规格书的要求和原则。

第 2 章

数据模型命名规则

2.1 建立命名规范的目的

建立命名规范的目的是为实体（表）、属性（列）建立有意义的、清晰的、简明的和独特的名字。一个完善的命名标准可以提高数据仓库的逻辑和物理模型的质量，同时也可以提高数据模型的易读和易懂性。

一个好的命名规范，应当使读者在读到一个列或表的名字时，可以立刻知道列或表的涵义，根据命名规范所创建的名字应当具有以下特点：

- 名字必须是一个有自我说明功能的标签。名字应该能够描述一个列或表中的内容。
- 列的名字应该能够告诉用户在列上可以进行哪些操作。

名字必须是独特的。实体/表的名字在一个企业的数据库模型中应该是唯一的，同样属性/列的名字应该在一个实体/表中是唯一的。

命名规范是一系列数据实体和数据元素的命名规则。通过遵守命名规范，数据模型设计师和数据库管理人员可以创造明确和独特的名字，并且避免不同的人取一些不一致甚至相互矛盾的名字，从而造成不必要的混乱。

2.2 命名规范说明

2.2.1 总体原则

- 为便于阅读和对照，逻辑模型中的实体和属性建议采用中文命名，物理模型中的表和字段采用拼音首字母命名，两者名字一一对应；具体情况视使用的工具而定。
- 物理模型表名限制在三十个字符之内。通常不要超过 15 个字符。
- 数据模型中所有的对象按照地震项目中的业务进行组织，类别代码如表 2.1。

表 2.1 业务类别命名代码表

业务类别	代码	说明
测震业务	CZ	
前兆业务	QZ	
信息业务	XX	
公共部分	CC	
强震动业务	QD	暂不涉及
活断层业务	DC	暂不涉及

2.2.2 数据模型版本命名规范

如果使用 PowerDesigner，最终的提交模型 PowerDesigner 模型文件采用如下形式：

数据模型+模型类型+版本号

其中：模型类型包括逻辑模型和物理模型两类。

版本号为 Vx.y.z 其中 x 为 Version 号，y 为 Release 号，z 为版本产出时间（年月日时分，如 200609010800，代表文档产出的日期为 2006 年 9 月 3 日 8 时 00 分）。

例如：地震监测数据模型_逻辑_V1.0.200609031400。

2.2.3 逻辑模型命名规则

1. 实体（Entity）命名规则

实体命名可以采用中文命名。视使用的建模工具定，如果使用 PowerDesigner 工具或 ERWin 工具，可以使用中文命名，如果是使用 Rational 工具，使用中文拼音代码代替。在注释中可以使用详细的中文名称。以下以 PowerDesigner 工具为例。

实体命名采用“实体限定词+实体类别名(+表)”的形式。实体限定词要采用标准的业务词汇。实体类别名用来描述表的内容的最基本的含义。比如：资料，类型，状态等。类别名为可选项。

2. 属性（Attribute）命名规则

属性命名采用中文命名。视使用的建模工具对中文的支持情况而定。

属性名在同一实体内唯一。

属性名应保持同名同义。

属性命名采用“属性限定词+属性类别名”的形式。属性类别名用来描述属性的内容的最基本的含义。比如：代码，名称，地址等。限定词用来更深入地阐明属性的含义和内容，能使一些意义含糊的名称变得定义更清晰。属性限定词要采用标准的业务词汇。

2.2.4 物理模型命名规则

1. 表（Table）命名规则

表名采用英文字母+数字命名。

表名采用“A_BB_CCCCCCCC”形式。

- A：代表对象类型（T 数据表，V 视图，I 索引）。
- BB：代表地震大的业务。
- CCCCCCCC：代表对象意义。与相对应的实体名一致，采用拼音首字母。

2. 字段（Column）命名规则

字段名采用英文字母+数字命名。

字段命名采用“限定词+_+类别名”的形式。与相对应的属性名一致，采用拼音首字母。