

Broadview®
www.broadview.com.cn

希赛®
CSAI .CN

秋季
最新
版



National Computer Rank Examination

全国计算机等级考试 考点分析、题解与模拟 四级数据库工程师

希赛IT教育研发中心 组编
钟经伟 唐强 主编

本书**关键字**：**紧扣大纲**、**真题分析**、**实战模拟**、**权威专家**、**辅导名师**、**在线考前模拟**、**平台强大**、**交流互动**、**过关无忧**



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>



National Computer Rank Examination

全国计算机等级考试 考点分析、题解与模拟 四级数据库工程师

希赛IT教育研发中心 组编
钟经纬 唐强 主编

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书由希赛 IT 教育研发中心组织编写,紧扣教育部考试中心新推出的考试大纲,通过对历年试题进行科学分析、研究、总结、提炼而成。本书主要内容包括应试指南、考点分析与同步训练、五套全真模拟试卷及解析,以及 2008 年 4 月、2008 年 9 月和 2009 年 3 月的真题解析。

基于历年试题,利用统计分析的方法,科学地做出结论并预测以后的出题动向,是本书的一大特色。本书可以保证既不漏掉考试必需的知识点,又不加重考生备考负担,使考生轻松、愉快地掌握知识点并领悟四级数据库工程师考试的真谛。“一书在手,过关无忧”。

本书可以作为参加全国计算机等级考试四级数据库工程师考试的复习用书,也可作为数据库系统工程师考试的复习用书,还可作为从事数据库设计与管理的人员学习参考用书。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

全国计算机等级考试考点分析、题解与模拟. 四级数据库工程师 / 钟经伟, 唐强主编; 希赛 IT 教育研发中心组编. —北京: 电子工业出版社, 2009.7

ISBN 978-7-121-08877-3

I. 全… II. ①钟…②唐…③希… III. ①电子计算机—水平考试—自学参考资料②数据库系统—水平考试—自学参考资料 IV. TP3

中国版本图书馆 CIP 数据核字 (2009) 第 079098 号

责任编辑: 葛 娜

印 刷: 北京智力达印刷有限公司

装 订: 北京中新伟业印刷有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 880×1230 1/16 印张: 26.75 字数: 878 千字

印 次: 2009 年 7 月第 1 次印刷

印 数: 4000 册 定价: 49.80 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。

质量投诉请发邮件至 zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线: (010) 88258888。

前言

现在的信息系统基本上都是基于数据库的系统，因此，在信息系统建设中，数据库设计和管理是很重要的工作，数据库工程师也是一个十分抢手的职业岗位。全国计算机等级考试（NCRE）由教育部考试中心主办，面向社会，是用于考查非计算机专业人员的计算机应用知识与能力的全国性计算机水平考试体系。该考试客观、公正，得到社会广泛认可。

内容超值，针对性强

本书紧扣教育部考试中心新推出的考试大纲，是为参加计算机等级考试四级数据库工程师的考生编写的，通过对历年试题进行科学分析、研究、总结、提炼而成。

本书着重对考试大纲的相关内容重点地细化和深化，并且提供真题分析，针对考点提供五套全真模拟试题，帮助考生掌握知识、把握重点，可以帮助考生掌握数据库技术的基本概念、原理、方法和技术；能够使用 SQL 语言实现数据库操作，具备数据库系统安装、配置及数据库管理与维护的基本技能；掌握数据库管理与维护、性能优化的基本方法；了解数据库应用系统的生命周期及其设计、开发过程；熟悉常用的数据库管理和开发工具，具备用指定的工具管理和开发简单数据库应用系统的能力；了解数据库技术的最新发展；最终满足数据库工程师岗位的要求。

作者权威，阵容强大

希赛是中国领先的 IT 教育和互联网技术公司，在 IT 人才培养、行业信息化、互联网服务及其他技术方面，希赛始终保持 IT 业界的领先地位。希赛对国家信息化建设和软件产业化发展具有强烈的使命感，利用希赛顾问网（www.CSAI.cn）强大的平台优势，加强与促进 IT 人士之间的信息交流和共享，实现 IT 价值。

希赛 IT 教育研发中心（以下简称“希赛教育”）是希赛公司属下的一个专门从事 IT 教育、教育产品开发、教育书籍编写的部门，在 IT 教育方面具有极高的权威性。特别是在 IT 在线教育方面，稳居国内首位，希赛教育的远程教育模式得到了国家教育部门的认可和推广。“让每个人随时随地享受 IT 教育”是希赛教育不懈努力和追求的目标。

希赛教育是国内唯一的进行计算机等级考试在线教育的大型教育机构，在该领域取得了很好的效果：

- 组织大纲制订者和阅卷组成员编写了考试辅导教材近 20 本，内容涵盖了计算机等级考试的二级、三级和四级的主要级别。
- 组织权威专家和辅导名师录制了考试培训视频教程，对历年考试进行了跟踪研究和比较研究，编写了权威的全真模拟试题。
- 希赛教育在线测试频道中有数十套全真模拟试题，供考生进行考前模拟，并且还开发了上机考试的模拟测试系统。
- 希赛教育的计算机等级考试培训采取统一教材、统一视频、统一认证教师的形式，采取线下培训与线上辅导相结合的方式，确保学员在通过考试的前提下能真正学到有用的知识。

本书由希赛 IT 教育研发中心组织编写，参加编写的人员来自大学教学一线和企业数据库设计与管理团队，具有丰富的教学和辅导经验，对等级考试有深入的研究，具有极强的应试技巧、理论知识、实践经验和责任心。

本书由钟经伟、唐强主编，张友生审核了所有稿件。全书共分 22 章，第 1 章由苏永乐编写，第 2 章

由吴蓉波编写，第 3、4、5 章由刘伟编写，第 6 章由刁洪祥编写，第 18、19、21、22 章由钟经伟和付毓谦编写，第 7、10、11、12、13 章由谢传银编写，第 8 章由刘伟和唐强编写，第 9 章由陈世帝编写，第 14、15 章由陈艳编写，第 20 章由程平编写，第 16、17 章由李宁编写，附录真题分析由刘建兵编写。

在线测试，心中有数

希赛教育网在线测试平台（test.educity.cn）为考生准备了在线测试，其中有数十套全真模拟试题和考前密卷，考生可选择任何一套进行测试。测试完毕，系统自动判卷，立即给出分数。

对于考生做错的地方，系统会自动记忆，待考生第二次参加测试时，可选择“试题复习”。这样，系统就会自动把考生原来做错的试题显示出来，供考生重新测试，以加强记忆。

如此，读者可利用希赛教育网在线测试平台的在线测试系统检查自己的实际水平，加强考前训练，做到心中有数，考试不慌。

诸多帮助，诚挚致谢

在本书出版之际，要特别感谢教育部考试中心计算机等级考试办公室的命题专家们，编者在本书中引用了部分考试原题，使本书能够尽量方便读者的阅读。在本书的编写过程中，参考了许多相关的文献和书籍，编者在此对这些参考文献的作者表示感谢。

感谢电子工业出版社孙学瑛老师，她在本书的策划、选题的申报、写作大纲的确定，以及编辑、出版等方面，付出了辛勤的劳动和智慧，给予了我们很多的支持和帮助。

感谢参加希赛教育计算机等级考试辅导和培训的学员，正是他们的想法汇成了本书的源动力，他们的意见使本书更加贴近读者。

由于编者水平有限，且本书涉及的内容很广，书中难免存在错漏和不妥之处，编者诚恳地期望各位专家和读者不吝指正和帮助，对此，我们将十分感激。

互动讨论，专家答疑

希赛教育网是中国最大的 IT 在线教育网站，是中国最大的 IT 资源平台，希赛教育网社区（bbs.educity.cn）是国内人气最旺的 IT 教育和技术社区。在希赛教育网社区“计算机等级考试”版块，读者可以和数十万考生进行在线交流，讨论有关学习和考试的问题，以及人生和职业规划的话题。希赛 IT 教育研发中心拥有强大的师资队伍，为读者提供全程的答疑服务，在线回答读者的问题。

有关本书的意见反馈和咨询，读者可在希赛教育网社区中的“书评在线”版块的“希赛 IT 教育研发中心”栏目上与作者进行交流。

希赛 IT 教育研发中心
2009 年 5 月

目 录

第 1 章 引论 1

1.1 基本概念	1
1.1.1 信息与数据	1
1.1.2 数据库系统	2
1.2 数据模型	2
1.2.1 考点 1: 数据模型的概念	3
1.2.2 考点 2: 数据模型的分类	3
1.3 数据视图和模式结构	7
1.3.1 考点 1: 三级模式结构	7
1.3.2 考点 2: 两级映像与数据独立性	9
1.4 数据库系统体系结构	10
1.5 数据库管理系统	12
1.5.1 考点 1: 数据库管理系统的主要功能	12
1.5.2 考点 2: DBMS 与周围系统硬件、软件的关系	13
1.5.3 考点 3: 数据库系统的全局结构	14
1.6 数据库技术的发展	15
1.7 同步训练	17
1.8 同步训练答案	18

第 2 章 数据库应用系统生命周期 19

2.1 数据库应用系统生命周期	19
2.1.1 考点 1: 软件工程与软件开发方法	19
2.1.2 考点 2: DBAS 软件组成	20
2.1.3 考点 3: DBAS 生命周期模型	21
2.2 规划与分析	22
2.2.1 考点 1: 系统规划与定义	22
2.2.2 考点 2: 可行性分析	22
2.2.3 考点 3: 项目规划	23

2.3 需求分析	24
2.3.1 考点 1: 数据与数据处理需求分析	24
2.3.2 考点 2: 业务规则需求分析	25
2.3.3 考点 3: 性能需求分析	26
2.3.4 考点 4: 其他需求分析	26
2.4 系统设计	27
2.4.1 考点 1: 概念设计	27
2.4.2 考点 2: 逻辑设计	28
2.4.3 考点 3: 物理设计	29
2.5 实现与部署	30
2.6 运行管理与维护	31
2.6.1 考点 1: 日常维护	31
2.6.2 考点 2: 系统性能监控和分析	32
2.6.3 考点 3: 系统性能优化调整	32
2.6.4 考点 4: 软件系统升级	33
2.7 同步训练	34
2.8 同步训练答案	34

第 3 章 需求分析与功能建模方法 35

3.1 需求分析概述	35
3.1.1 考点 1: 需求分析概念和任务	35
3.1.2 考点 2: 需求来源与需求获取方法	36
3.1.3 考点 3: 需求分析的过程	37
3.1.4 考点 4: 软件需求规格说明书	38
3.2 DFD 建模方法	38
3.2.1 考点 1: DFD 基本图形符号	38
3.2.2 考点 2: 绘制数据流图	40
3.2.3 考点 3: 数据字典和加工说明	42
3.3 IDEF0 建模方法	43

3.3.1	考点 1: IDEF0 方法	43
3.3.2	考点 2: IDEF0 建模过程	45
3.4	DFD 与 IDEF0 的比较	46
	考点: DFD 与 IDEF0 的比较	46
3.5	同步训练	46
3.6	同步训练答案	48

第 4 章 数据库概念设计及数据建模.....50

4.1	数据库概念设计概述	50
4.1.1	考点 1: 概念设计的任务	50
4.1.2	考点 2: 概念设计方法和步骤	51
4.2	数据建模方法	52
	考点: 数据建模方法	52
4.3	E-R 建模方法	52
4.3.1	考点 1: 基本概念	52
4.3.2	考点 2: E-R 图	53
4.3.3	考点 3: E-R 建模步骤	54
4.3.4	考点 4: E-R 建模实例	55
4.4	IDEF1X 建模方法	60
4.4.1	考点 1: IDEF1X 模型元素	60
4.4.2	考点 2: IDEF1X 建模过程	62
4.5	同步训练	64
4.6	同步训练答案	67

第 5 章 关系数据库逻辑设计.....69

5.1	概述	69
5.2	基本概念	70
5.2.1	考点 1: 基本术语	70
5.2.2	考点 2: 关系的完整性	71
5.2.3	考点 3: 关系代数	71
5.2.4	考点 4: 关系数据库	75
5.3	关系数据库设计理论	76
5.3.1	考点 1: 问题提出	76
5.3.2	考点 2: 函数依赖	76
5.3.3	考点 3: 规范化设计方法	79
5.4	数据库模式设计	82
5.4.1	考点 1: 初始关系模式设计	82
5.4.2	考点 2: 优化关系模式	86
5.4.3	考点 3: 数据完整性设计及 外模式设计	90
5.5	同步训练	90
5.6	同步训练答案	96

第 6 章 存储技术与数据库物理设计.....100

6.1	文件组织	100
6.1.1	考点 1: 定长记录表示法	100
6.1.2	考点 2: 变长记录表示法	101
6.2	文件结构与存取	102
6.2.1	考点 1: 堆文件	102
6.2.2	考点 2: 顺序文件	103
6.2.3	考点 3: 多表聚集文件	103
6.2.4	考点 4: 索引文件	105
6.2.5	考点 5: 散列文件	105
6.3	索引技术	105
	考点: 索引的一些基本概念	105
6.4	散列技术	109
6.4.1	考点 1: 散列文件的操作	109
6.4.2	考点 2: 散列函数	110
6.4.3	考点 3: 桶溢出控制	110
6.5	数据字典	111
6.6	数据库物理设计	112
6.6.1	考点 1: 设计内容和目标	112
6.6.2	考点 2: 确定数据的存储结构	112
6.6.3	考点 3: 选择和设计存取路径	113
6.6.4	考点 4: 数据分布设计	114
6.6.5	考点 5: 确定系统配置	115
6.6.6	考点 6: 物理模式评估	116
6.7	同步训练	117
6.8	同步训练答案	118

第 7 章 数据库应用系统功能设计.....120

7.1	软件体系结构与设计过程	120
7.1.1	考点 1: 软件体系结构	120
7.1.2	考点 2: 软件设计过程	121
7.2	DBAS 总体设计	121
7.2.1	考点 1: 系统总体设计步骤	122
7.2.2	考点 2: 结构化设计原理	123
7.2.3	考点 3: C/S 和 B/S 体系结构	123
7.3	概要设计	124
7.3.1	考点 1: 数据库应用系统的 概要设计	125
7.3.2	考点 2: 数据抽象与局部视图 设计	125
7.3.3	考点 3: 视图的集成	125
7.3.4	考点 4: 概要设计中模块化 原则	126

7.4	详细设计.....	127
	考点: 数据库事务详细设计	127
7.5	人机界面设计.....	128
7.6	同步训练	129
7.7	同步训练答案	131

第 8 章 关系数据库操作语言 132

8.1	SQL 支持的数据类型	132
	考点: SQL Server 数据类型	132
8.2	定义和维护关系表	134
	8.2.1 考点 1: 创建关系表	134
	8.2.2 考点 2: 修改和删除关系表	137
8.3	数据操作语言	139
	8.3.1 考点 1: 查询数据	139
	8.3.2 考点 2: 插入数据	148
	8.3.3 考点 3: 更新数据	149
	8.3.4 考点 4: 删除数据	150
8.4	索引	151
	考点: 创建和删除索引	151
8.5	视图	153
	8.5.1 考点 1: 创建视图	153
	8.5.2 考点 2: 修改和删除视图	154
	8.5.3 考点 3: 视图的作用	155
8.6	同步训练	155
8.7	同步训练答案	161

第 9 章 事务调度与并发控制 166

9.1	事务的基本知识	166
	9.1.1 考点 1: 概念	166
	9.1.2 考点 2: 特征	168
	9.1.3 考点 3: 事务调度	168
	9.1.4 考点 4: 并发执行的正确性 问题	170
9.2	并发控制技术	174
	9.2.1 考点 1: 基于锁的并发控制	174
	9.2.2 考点 2: 封锁粒度	175
	9.2.3 考点 3: 封锁协议	176
	9.2.4 考点 4: 两段锁协议	176
9.3	死锁与活锁	178
	9.3.1 考点 1: 死锁	178
	9.3.2 考点 2: 活锁	179
9.4	同步训练	180

9.5	同步训练答案	182
-----	--------------	-----

第 10 章 数据库的实施、运行和维护 183

10.1	数据库的实施	183
	10.1.1 考点 1: 数据库的实施	183
	10.1.2 考点 2: 定义数据库结构	183
	10.1.3 考点 3: 数据装载	184
	10.1.4 考点 4: 编制与调试应用程序	184
	10.1.5 考点 5: 试运行工作的内容	184
10.2	数据库的运行和维护	185
10.3	监控分析	186
10.4	空间管理	187
	10.4.1 考点 1: 数据库空间管理	188
	10.4.2 考点 2: 确定数据库的数目	188
10.5	参数调整	189
10.6	查询优化	190
	10.6.1 考点 1: 查询优化	190
	10.6.2 考点 2: 查询优化步骤	192
10.7	同步训练	193
10.8	同步训练答案	194

第 11 章 故障管理 195

11.1	事务	195
11.2	故障的种类及解决方法	197
	11.2.1 考点 1: 事务内部故障	197
	11.2.2 考点 2: 系统故障	197
	11.2.3 考点 3: 介质故障	198
	11.2.4 考点 4: 计算机病毒故障	198
11.3	数据库恢复技术概述	199
11.4	数据转储	199
11.5	登记日志文件	201
	11.5.1 考点 1: 日志文件的格式 和内容	201
	11.5.2 考点 2: 日志文件的作用	201
	11.5.3 考点 3: 登记日志文件的原则	202
11.6	具有检查点的恢复技术	203
	11.6.1 考点 1: 检查点的作用	203
	11.6.2 考点 2: 检查点的引入	203
	11.6.3 考点 3: 恢复的步骤	204
11.7	数据库镜像	204
	11.7.1 考点 1: 数据库镜像的引入	204
	11.7.2 考点 2: 数据库镜像简介	205

11.7.3	考点 3: 数据库镜像的分类	206
11.7.4	考点 4: 数据库镜像的工作方式	207
11.7.5	考点 5: 运行模式	207
11.8	RAID 的恢复技术	207
11.9	同步训练	208
11.10	同步训练答案	211

第 12 章 SQL Server 数据库管理系统215

12.1	SQL Server 2005 概述	215
12.2	SQL Server 2005 的安装	216
12.2.1	考点 1: 安装前的准备	217
12.2.2	考点 2: SQL Server 升级	217
12.3	创建与管理数据库	219
12.4	Transact-SQL 简介	220
12.5	数据传输	223
12.6	同步训练	224
12.7	同步训练答案	225

第 13 章 数据库对象277

13.1	存储过程	227
13.1.1	考点 1: 存储过程的基本概念	227
13.1.2	考点 2: 创建和执行存储过程	228
13.2	用户自定义函数	228
13.2.1	考点 1: 基本概念	228
13.2.2	考点 2: 三种函数的特点	229
13.2.3	考点 3: 修改和删除用户 自定义函数	229
13.3	触发器	229
13.3.1	考点 1: 触发器的基本概念	229
13.3.2	考点 2: 创建触发器	230
13.3.3	考点 3: 修改触发器	231
13.3.4	考点 4: 删除触发器	231
13.4	查看、修改及删除对象	231
13.4.1	考点 1: 查看对象	231
13.4.2	考点 2: 修改对象	232
13.4.3	考点 3: 删除对象	232
13.5	同步训练	232
13.6	同步训练答案	234

第 14 章 安全管理235

14.1	安全控制	235
14.1.1	考点 1: 安全控制模型	235

14.1.2	考点 2: 数据库权限的种类及 用户的分类	236
--------	--------------------------------	-----

14.2 SQL Server 的安全控制 236

14.3 管理 SQL Server 登录账号 237

14.3.1	考点 1: 创建登录账号	237
14.3.2	考点 2: 删除登录账号	238

14.4 管理数据库用户 239

14.4.1	考点 1: 建立数据库用户	239
14.4.2	考点 2: 删除数据库用户	240

14.5 用户管理权限 241

14.5.1	考点 1: 用户权限类型	241
14.5.2	考点 2: 用户权限管理	241

14.6 角色 244

14.6.1	考点 1: 服务器角色	244
14.6.2	考点 2: 数据库角色	245

14.7 同步训练 249

14.8 同步训练答案 249

第 15 章 备份与恢复数据库250

15.1 备份数据库 250

15.1.1	考点 1: 备份的概念与方式	250
15.1.2	考点 2: 备份工作的规划	251
15.1.3	考点 3: 备份设备	251
15.1.4	考点 4: 备份的执行	252

15.2 恢复数据库 255

15.2.1	考点 1: 恢复的概念与方式	255
15.2.2	考点 2: 利用对象资源管理器 恢复数据库	256
15.2.3	考点 3: 使用 Transact-SQL 语句 恢复数据库	258

15.3 同步训练 259

15.4 同步训练答案 260

第 16 章 VB 开发环境和数据访问 接口261

16.1 Visual Basic 6.0 简介 261

16.1.1	考点 1: 什么是 Visual Basic 6.0	261
16.1.2	考点 2: Visual Basic 6.0 的版本 和软、硬件要求	263

16.2 Visual Basic 程序设计基础 263

16.2.1	考点 1: Visual Basic 6.0 的集成开发环境	263
16.2.2	考点 2: 编写 Visual Basic 6.0 程序的过程	264
16.3	数据绑定控件	264
16.3.1	考点 1: 数据绑定概述	265
16.3.2	考点 2: 单表数据绑定控件	266
16.3.3	考点 3: DBList 和 DBCombo 控件	267
16.4	数据库访问接口	268
16.4.1	考点 1: 什么是数据库访问接口	269
16.4.2	考点 2: ODBC	270
16.4.3	考点 3: OLE DB	272
16.4.4	考点 4: ADO	272
16.5	同步训练	273
16.6	同步训练答案	274

第 17 章 VB 数据库应用程序 275

17.1	ADO 数据控件	275
17.1.1	考点 1: 什么是 ADO 数据控件	275
17.1.2	考点 2: 连接数据源的方式	276
17.1.3	考点 3: ADO 数据控件的主要属性	277
17.2	ADO 对象	278
17.2.1	考点 1: Connection 对象	278
17.2.2	考点 2: Recordset 对象	280
17.2.3	考点 3: Command 对象	283
17.2.4	考点 4: Parameter 对象	285
17.2.5	考点 5: Field 对象	285
17.3	同步训练	285
17.4	同步训练答案	286

第 18 章 统一建模语言 287

18.1	UML 简介	287
18.2	UML 静态建模机制	289
18.2.1	考点 1: 用例图	289
18.2.2	考点 2: 类图和对象图	291
18.2.3	考点 3: 包	294
18.2.4	考点 4: 构件图和配置图	294
18.3	UML 动态建模机制	296

18.3.1	考点 1: 序列图	296
18.3.2	考点 2: 协作图	297
18.3.3	考点 3: 状态图	298
18.3.4	考点 4: 活动图	299
18.4	同步训练	301
18.5	同步训练答案	302

第 19 章 数据库新技术 304

19.1	分布式数据库	304
19.1.1	考点 1: 分布式数据库系统概述	304
19.1.2	考点 2: 分布式数据库的分类	305
19.1.3	考点 3: 分布式数据库的特点和优缺点	305
19.1.4	考点 4: 数据分片	306
19.1.5	考点 5: 数据分配方式	306
19.1.6	考点 6: 分布式数据库系统模式结构	307
19.1.7	考点 7: 分布式数据库管理系统功能结构	307
19.1.8	考点 8: 分布式数据库系统查询处理和优化	308
19.1.9	考点 9: 分布式事务管理	309
19.2	对象数据库	311
19.2.1	考点 1: 面向对象数据模型	311
19.2.2	考点 1: 对象-关系数据库	312
19.3	并行数据库	314
19.3.1	考点 1: 并行数据库的并行结构	315
19.3.2	考点 2: 数据划分	315
19.3.3	考点 3: 并行操作算法和查询优化	315
19.4	同步训练	317
19.5	同步训练答案	319

第 20 章 数据仓库和数据挖掘 321

20.1	决策支持系统的发展	321
20.1.1	考点 1: 决策支持系统的产生与发展	321
20.1.2	考点 2: 数据仓库与决策支持系统	322
20.2	数据仓库技术概述	324

20.2.1	考点 1: 数据仓库的定义 和特点	324
20.2.2	考点 2: 数据仓库的体系结构	325
20.2.3	考点 3: ETL 介绍	326
20.2.4	考点 4: 元数据	326
20.2.5	考点 5: 操作型数据存储	326
20.3	设计和构建数据仓库	327
20.3.1	考点 1: 数据仓库的设计	327
20.3.2	考点 2: 数据仓库的数据模型	328
20.3.3	考点 3: 数据仓库的构建步骤	328
20.3.4	考点 4: 两个重要的设计因素	329
20.4	数据仓库的运行与维护	330
20.4.1	考点 1: 数据仓库数据的更新 维护	330
20.4.2	考点 2: 数据仓库监控与 元数据管理	331
20.5	联机分析处理与多维数据模型	331
20.5.1	考点 1: 联机分析处理的概念	331
20.5.2	考点 2: 多维基本概念	331
20.5.3	考点 3: 多维分析	331
20.5.4	考点 4: 联机分析处理的 实现方式	332
20.6	数据挖掘技术	332
20.6.1	考点 1: 知识发现过程	333
20.6.2	考点 2: 关联规则挖掘	333
20.6.3	考点 3: 分类挖掘	333
20.6.4	考点 4: 聚类挖掘	334
20.6.5	考点 5: 时间序列分析	334
20.7	同步训练	335
20.8	同步训练答案	336

第 21 章 模拟试卷 337

21.1	模拟试卷一	337
21.2	模拟试卷二	342
21.3	模拟试卷三	346
21.4	模拟试卷四	350
21.5	模拟试卷五	354

第 22 章 模拟试卷解析与答案 359

22.1	模拟试卷一解析与答案	359
22.2	模拟试卷二解析与答案	362
22.3	模拟试卷三解析与答案	366
22.4	模拟试卷四解析与答案	371
22.5	模拟试卷五解析与答案	375

附录 A 2008 年上半年笔试试卷及 解析 379

附录 B 2008 年下半年笔试试卷及 解析 388

附录 C 2009 年上半年笔试试卷及 解析 399

附录 D 计算机等级考试四级数据库 工程师考试大纲 413

参考文献 415

引 论

20 世纪 60 年代末, 数据库技术就是作为数据管理中的一项新技术发展起来的。它是计算机软件领域的一个重要分支, 经过近 50 年的发展, 形成了较为完整的理论体系和实用技术。在系统地介绍数据库技术前, 本章首先介绍主要的基本概念、基本原理及一些重要术语, 以及数据库技术的发展和应用情况, 为后续各章奠定理论基础。在历年的考试中, 虽然直接针对本章考题的分值不大, 一般占 4 分左右, 但本章是全书的基础, 只有在深刻领会本章的基本概念及术语的基础上, 重点掌握好本章的主要考点, 才可能领会其他各章考点的内容。

本章重要考点主要有:

- 数据模型及其分类、数据库系统三级模式结构
- 数据库管理系统的功能、结构及组成
- 数据库系统全局结构

1.1 基本概念

本节介绍的基本概念和术语是全书的基础, 应深刻领会, 以便对数据库系统的研究对象及主要内容有初步的了解, 为掌握后续章节考点打下基础。

1.1.1 信息与数据

信息 (information) 是现实世界在人脑中的抽象反映, 是通过人的感官感知出来并经过人脑的加工而形成的反映现实世界中事物的概念, 是现实世界事物的存在方式和运动状态的反映。信息有许多重要的特征:

- 信息来源于物质和能量
- 信息是可以感知的
- 信息是可以存储的
- 信息是可以加工、传递和再生的

这些特征, 构成了信息的最重要的自然属性。作为信息的社会属性, 信息已经成为社会上各行各业不可缺少的重要资源之一。人类获取、积累并利用信息是认识和改造客观世界的必要过程。

数据 (data) 是一切文字、符号、声音、图像等有意义的组合。它是描述现实世界中各种具体事物或抽象概念的可存储并具有明确意义的信息; 它是用符号记录下来的、可以识别的信息; 它是信息的符号表

示或称为载体。

数据是数据库系统研究和处理的对象。数据和信息是分不开的，它们之间既有联系又有区别。信息是以数据的形式表示的，即它是以数据为载体而表现的。信息是数据的内涵，是对数据语义的解析；而数据则是表示信息的一种手段，是信息的载体。数据表示了信息，而信息只有通过数据形式表示出来才能被人们理解和接受。

数据处理（data processing）又称为信息处理，是指从某些已知的表示某些信息的数据出发，推导整理出一些新的数据，从而又表示出一些新的信息的过程。它涉及数据的收集、管理、加工直至产生新信息输出的全过程。

数据管理（data management）是指数据的收集、整理、组织、存储、维护、检索及传送等操作处理过程。

数据处理和数据管理是互相联系的，数据管理中的各种操作都是数据处理业务的必不可少的基本环节，数据管理技术的好坏，直接影响到数据处理的效率。数据管理是数据处理的中心问题。数据处理的目的是有两个：一是借助计算机科学地保存和管理大量复杂的数据，以便人们能方便而充分地利用这些宝贵的信息资源；二是从大量的原始数据中抽取、推导出对人们有价值的信息。

1.1.2 数据库系统

数据库（DataBase，DB）是指相互关联的数据集合。或者说数据库是长期存储在计算机内、有组织的、统一管理的数据集合。数据库中的数据按一定的数据规范进行组织、描述和存储，为各种用户共享。数据库中的数据应具有最小冗余度、数据间的联系密切、对程序又具有较高独立性等特点。

在应用领域中，存在着众多复杂的各类数据。为了在数据库中能科学合理地组织和存储这些数据，需要利用数据库管理系统这一重要的数据管理软件。

数据库管理系统（DataBase Management System，DBMS）是数据库系统的核心，是用于建立、使用和维护数据库的一组软件。一般情况下，数据库管理系统是建立在操作系统的基础之上，位于操作系统与用户之间的一层数据管理软件，负责对数据库进行统一的管理和控制。

数据库管理系统给用户提供了一个软件环境，对外它向用户提供一整套命令，利用这些命令，用户可以快速方便地建立、维护、检索、存取和处理数据库中的信息，包括 DB 的建立、查询、更新及各种数据控制等。数据库管理系统的主要目标，是使数据作为一种可管理的资源来处理。数据库管理系统分为多个程序模块，每个模块实现数据库系统的一种功能。数据库管理系统不仅承担执行各种应用程序对数据库中数据的操作命令，还要承担数据库的维护工作，以保证数据库的安全性和完整性。数据库管理系统总是基于某种数据模型，可以分为层次模型、网状模型、关系模型和面向对象模型等。有关数据模型的概念，将在 1.2 节中介绍，并在 1.5 节中进一步阐述数据库管理系统的功能、工作原理和全局结构。

数据库系统（DataBase System，DBS）是采用了数据库技术的计算机系统，是具有管理和控制数据库功能的计算机系统。它是在计算机系统中引入数据库技术后由硬件、软件构成的系统。DBS 可分为硬件、软件 and 用户 3 个层次。硬件是指计算机硬件平台，由于存储数据的容量很大，数据库系统要求计算机硬件平台具有较快的 CPU、足够大的内存和外存、较高的系统通信能力；软件包括系统软件和应用软件，诸如 DBMS、支持 DBMS 运行的操作系统、数据库应用开发工具和数据库应用程序等；用户包括数据库管理员（DataBase Administrator，DBA）、系统分析员、应用程序员和终端用户等。

1.2 数据模型

数据模型是数据库系统的核心和基础，也是本章的重要考点之一。

1.2.1 考点 1: 数据模型的概念

模型是对现实世界的抽象。在数据库技术中我们使用模型来描述数据库的结构与语义。一个具体的数据库是针对某个应用领域应用数据的集合,它反映了应用数据本身的内容及相互之间的联系。由于计算机无法直接处理现实世界中的具体事物,因此我们必须将具体事物转换成计算机能够处理的数据。为此,在数据库中使用数据模型这个工具来抽象、表示和处理现实世界中的数据和信息,以便为计算机所识别。

通常把表示实体类型及实体之间联系的模型称为数据模型。

数据模型是数据库系统的核心和基础,是用来描述数据的一组概念和定义,包括描述数据、数据联系、数据操作、数据语义及数据一致性等。数据模型给出了数据库系统中数据组织、存储和管理必须遵循的规范,这些规范精确地描述了系统的静态特征、动态特征和完整性约束条件。

数据模型一般应满足以下 3 方面的要求:

- (1) 能比较真实地模拟现实世界;
- (2) 容易为人们理解;
- (3) 便于在计算机上实现。

目前,还没有一个数据模型能够同时满足这 3 方面的要求。在数据库系统中,我们将针对不同的抽象层次、不同应用领域的特征和使用对象采用不同的数据模型。

数据模型的抽象层次如图 1-1 所示。

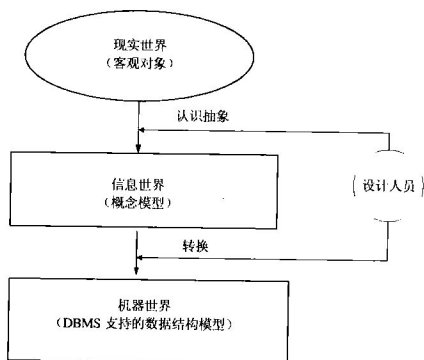


图 1-1 数据模型的抽象层次

例 1.2.1 对现实世界进行第一层抽象的模型是 ()。

- A. 实体-联系模型 B. 关系模型 C. 数据结构模型 D. 概念模型

解析: 设计人员首先从数据建模的角度将现实世界中的客观数据对象抽象组织为信息世界中的一种不依赖于具体计算机系统和 DBMS 的信息结构,人们称其为概念数据模型或信息模型。这是对现实世界进行的第一层抽象。

答案: D

例 1.2.2 对现实世界进行第二层抽象的模型是 ()。

- A. 实体-联系模型 B. 关系模型 C. 数据结构模型 D. 概念模型

解析: 设计人员要将对现实世界进行第一层抽象后的结果转换为机器世界中某一具体 DBMS 能支持的数据模型,人们称其为数据结构模型,并存储在计算机系统中。这是对现实世界进行的第二层抽象。

答案: C

1.2.2 考点 2: 数据模型的分类

根据模型应用目的的不同,可以将数据模型划分为 3 类,分别表示不同层次应用目的:第一类模型是概念数据模型,也称为信息模型;第二类模型是数据结构模型;第三类模型是物理数据模型。

1. 概念数据模型

概念数据模型 (conceptual data models) 也称为信息模型, 它是信息世界中的数据模型。概念数据模型按照用户的观点来对数据和信息进行建模, 它主要用于数据库设计, 它是作为从现实世界到其他数据模型转换的中间模型。概念模型是对现实世界的第一层抽象。概念模型不考虑数据的操作, 而只是用比较有效、自然的方式描述现实世界的数据及其联系。

概念模型不仅是数据库设计人员对现实世界研究的产物, 更为重要的是, 它是数据库设计人员与用户进行交流的工具。概念模型是从用户的观点对现实世界的抽象。因此, 概念模型不仅能完整地表现设计人员的思想, 而且简单清晰, 容易为人们所理解。

在设计概念模型时, 最著名、最实用的是 P.P.S.Chen 于 1976 年提出的“实体-联系模型”(Entity-Relationship Approach, E-R 模型)。E-R 模型是概念设计的重要工具, 它是基于对现实世界的这样一种认识: 世界由一组称做实体的基本对象以及这些对象间的联系组成。

E-R 模型是一种语义模型, 模型的语义方面主要体现在模型力图去表达数据的意义。E-R 模型在将现实世界中事物的含义和相互关联映射到概念模式方面是非常有用的。用 E-R 数据模型对一个系统的模拟, 称为 E-R 数据模式。E-R 数据模式可以很方便地转换成相应的关系、层次和网状数据模式。因此, 许多数据库设计工具都利用了 E-R 模型的概念。

E-R 模型采用实体 (entity)、实体集 (entity set)、联系 (relationship)、联系集 (relationship set) 和属性 (attribute) 等基本元素, 抽象描述现实世界中客观数据对象及其特征和数据对象之间的关联关系。

(1) 实体。实体是客观事物的反映, 它是现实世界中客观存在并可独立区别于其他对象的一个“事件”或“物体”。实体既可以是实际存在的对象, 也可以是某种概念, 例如, 一个工厂、一个车间、一种操作流程等。实体必须可相互区分。

(2) 属性。事物是有特性的, 反映在实体上, 就是实体的属性。属性描述了实体和实体集所具有的特征。一个实体具有有限个属性, 也可以说是这些属性的总和组成了这个实体。任一实体在任一属性上的取值都是有限制的, 一个属性的取值范围就是这个属性的域 (domain), 每个属性都有一个值域。例如, 姓名属性的域可定为 4 个汉字长的字符串, 职工号可定为 7 位整数等。

(3) 实体集。实体集是具有相同类型及相同性质 (属性) 的实体集合, 即所有属性名完全相同的实体集合在一起, 称为实体集。例如, 一名具体职工是实体, 全体职工就是一个实体集。能唯一地标识每个实体的属性或属性集称为键 (key) 或实体标识符。例如, 对职工实体集, 职工号是键。

(4) 联系。联系是实体之间的相互关系。联系分为实体内部的联系和实体与实体之间的联系两种。实体内部的联系描述了一个实体各属性间的关联关系; 实体之间的联系反映了现实世界中客观数据对象间的关联关系。联系类型还可以发生在多于两个的实体类型之间, 也可以发生在同一个实体类型之间。两个实体间的联系有一对一联系 (1:1)、一对多联系 (1:n) 和多对多联系 (m:n) 等种类。联系集 (relationship set) 是指同类联系的集合。联系和联系集也可具有属性, 同样可以用来描述联系和联系集的特征。

表示 E-R 模型的图称为 E-R 图。在 E-R 图中, 用矩形框表示实体类型; 用菱形表示联系类型; 用椭圆框表示实体类型和联系类型的属性; 相应的命名计入框内; 对键属性, 在属性名下画一条横线。

例如, 图 1-2 和图 1-3 分别表示职工实体集和学生选课 E-R 图。

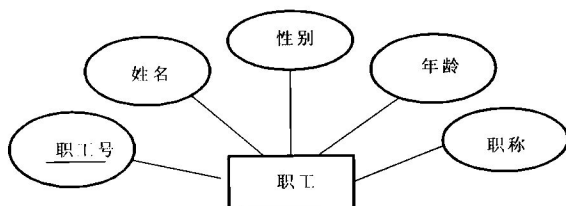


图 1-2 职工实体的 E-R 图表示

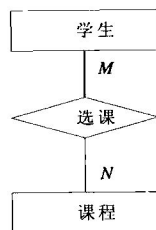


图 1-3 学生选课 E-R 图表示

如图 1-3 所示, 在学生选课中, 由于学生可选多门课程, 每门课程可以有多个学生选, 所以学生和课程之间关系是多对多的关系。

由于 E-R 图接近于人的思维, 容易理解; 而且, 由于 E-R 图与计算机无关, 用户容易接受, 因此 E-R 图可作为设计者与用户交流的重要工具。现在, E-R 图已成为软件工程的一个重要设计方法。

E-R 模型及建模方法将在第 4 章中进一步做详细阐述。

2. 数据结构模型

数据结构模型是直接面向数据库的逻辑结构, 是对现实世界的第二层抽象。它是按照计算机系统的观点对数据进行建模, 主要用于 DBMS 的实现。平时所说的数据模型, 主要是指数据结构模型, 它是数据库系统的核心和基础, 任何 DBMS 都支持一种数据结构模型。任何一种数据结构模型都规定了一种数据结构, 它给出了信息世界中实体和实体间联系的表示方法。数据结构模型描述了系统的静态特性, 这是数据模型最本质的内容。

数据结构模型的严格定义应包括 3 部分内容:

(1) 数据结构。数据结构用于描述系统静态特征。它是对实体类型和实体间联系的表达和实现; 它从语法的角度表示了现实世界中数据对象本身的结果和数据对象间的联系。通常按照数据结构的类型来命名各种不同的数据模型。例如, 层次结构、网状结构和关系结构分别被命名为层次模型、网状模型和关系模型等。

(2) 数据操作。数据操作用于描述系统的动态特征。表示对数据库的检索和更新 (包括插入、删除和修改) 两类操作。

(3) 数据完整性约束。数据完整性约束给出数据及其联系所具有的制约和依赖规则, 是一组完整性规则的集合。

实际的数据库管理系统所支持的主要数据结构模型有: 层次模型 (hierarchical model)、网状模型 (network model)、关系模型 (relational model) 和面向对象模型 (object oriented model)。

(1) 层次模型。用树型 (层次) 结构表示实体类型及实体间联系的数据模型称为层次模型。层次结构是树型结构, 树的结点是记录类型, 非根结点有且只有一个父结点。上层记录类型和下层记录类型的联系是 1:N 联系。

层次结构的特点是记录之间的联系通过指针来实现, 查询效率较高。但它存在只能表示 1:N 联系, 且查询和更新操作复杂, 应用程序的编写也比较复杂等缺点。

层次模型的典型实例是 1968 年美国 IBM 推出的 IMS 系统, 这个系统在 20 世纪 70 年代有着广泛的应用。

(2) 网状模型。用有向图 (网状) 结构表示实体类型及实体间联系的数据模型称为网状模型。有向图中的结点是记录类型, 箭头表示从箭头的记录类型到箭尾的记录类型的联系是 1:N 的联系。

网状模型的特点是记录之间的联系通过指针来实现, 网状模型的查询效率较高, 对 M:N 联系也容易实现。但存在编程比较复杂, 程序员必须在了解数据库的逻辑结构后才能进行编程等缺点。

20 世纪 70 年代的 DBMS 产品大部分都是基于网状模型的。

(3) 关系模型。关系模型是目前最重要的一种数据模型。关系数据模型是以集合论中的关系概念为基础发展起来的数据模型, 而关系运算是关系数据模型的理论基础。关系模型是数学化的模型, 将集合论、数理逻辑等知识引入关系模型中。

关系模型是用二维表格表示实体及实体间的联系。用表格结构表达实体集是关系模型的主要特征。从用户的角度看, 关系是一张表格, 表中的行对应于数据记录, 表中的列对应于数据记录的属性, 记录之间的联系是通过模式的键来体现的。在关系模型中, 基本的数据结构是表格, 整个数据库由多张表格组成。

关系模式相当于记录类型, 它的实例是一张表格, 称为关系。关系模型是由若干个关系模式组成的集合。数据库设计者可以从 E-R 图出发, 将 E-R 图转换为一系列关系表, 这些关系表组成了关系数据库。

关系模型和层次模型、网状模型的最大差别是用键而不是用指针导航数据，其表格简单，用户易懂，只需要用简单的查询语句就可以对数据库进行操作，并不涉及存储结构、访问技术等细节。

关系数据库系统就是采用关系模型作为数据的组织方式。基于关系模型的关系数据库系统是目前市场上的主流产品。关系数据库系统的典型实例有 SystemR、SQL Server、DB2、Oracle、Ingress、Sysbase、Informix、dBASE、FoxBase、FoxPro 等。

关系模型及关系数据库的设计在第 5 章中做详细阐述。

(4) 面向对象模型。面向对象模型是一种可扩充的数据模型，它吸收了语义数据模型和知识表示模型的一些基本概念，同时又借鉴了面向对象程序设计语言和抽象数据类型的一些思想。面向对象模型中的最基本概念是对象 (object) 和类 (class)。对象是现实世界中实体的模型化，它与记录概念相仿，但远比记录复杂。每一个对象都有一个唯一的标识符，把状态和行为封装在一起。其中，对象的状态是该对象属性值的集合；对象的行为是在对象状态上操作的方法集。共享同一属性集和方法集的所有对象构成一个类，一个类从类层次中直接或间接超类那里继承所有的属性和方法。用这个方法实现软件的可重用性。

3. 物理数据模型

物理数据模型属于底层的数据模型，主要用于实现对物理数据的存储和检索，它反映了数据库的物理结构。

物理数据模型用于描述数据库的存储记录格式、存储记录安排和存取方法。在数据库中，存取路径是一个特殊的结构，用于在数据库文件中有效地搜索一个特定的数据库记录。显然，数据库的物理结构设计完全依赖于给定的硬件环境和数据库产品。

 **例 1.2.3** 数据模型可以分为 3 类，正确的是 ()。

- A. E-R 模型、数据结构模型、物理数据模型
- B. 概念数据模型、关系模型、面向对象模型
- C. 概念数据模型、数据结构模型、物理数据模型
- D. 层次模型、网状模型、关系模型

解析：数据模型的分类是根据模型应用的不同目的、不同的应用层次而划分的。E-R 模型是作为一种数据库设计方法提出来的，所以它不属于数据模型分类的对象，因此 A 选项是不正确的。关系模型、层次模型、网状模型和面向对象模型，它们是属于同一个应用层次但数据结构不同的数据模型，所以 B 和 D 选项都是不正确的。数据模型分类的方法和数据结构模型分类的方法是不同的，不能混在一起。

答案：C

 **例 1.2.4** 下列选项中全不属于数据结构模型的是 ()。

- A. E-R 模型、信息模型、实体-联系模型
- B. 概念数据模型、关系模型、面向对象模型
- C. 层次模型、E-R 模型、实体-联系模型
- D. 层次模型、网状模型、关系模型、面向对象模型

解析：数据结构模型是从数据管理系统的角度，按照数据结构的特征而建立的数据模型，每一种数据结构模型都应是某一个实际的数据库管理系统所能支持的一种数据模型。因此，凡是不针对 DBMS 所支持的数据模型，都不应属于数据结构模型。由于层次模型、网状模型、关系模型和面向对象模型都属于数据结构模型，所以 B、C、D 选项都是不正确的。只有 A 选项中所列的模型均是不针对具体的 DBMS 所支持的数据结构。不过应注意，在不引起混淆的情况下，通常所说的数据模型指的是数据结构模型。

答案：A

 **例 1.2.5** 数据模型是用来表示实体及实体间联系的。网状模型、层次模型和关系模型都是数据库中的基本数据模型。在实体及实体间联系的表示方法上，网状模型可采用 (1)，层次模型可采用 (2)，关系模型则采用 (3)。