

Y 全国税务系统岗位专业知识与技能培训系列教材

PRACTICAL KNOWLEDGE

AND SKILLS OF TAXATION 国家税务总局教材编写组 编

数据库 与数据仓库



人 大 社

中国图书馆学会图书馆分会图书馆学情报学专业委员会

PRACTICAL KNOWLEDGE

AS SOURCE OF DATA

知识库 与数据仓库



中国图书馆学会图书馆分会图书馆学情报学专业委员会

全国税务系统岗位专业知识与技能培训系列教材

数据库与数据仓库

国家税务总局教材编写组 编

人 民 出 版 社

责任编辑:李春生

装帧设计:肖 辉

图书在版编目(CIP)数据

数据库与数据仓库/国家税务总局教材编写组编.

-北京:人民出版社,2004.8

(全国税务系统岗位专业知识与技能培训系列教材)

ISBN 7-01-004489-9

I.数… II.国… III.数据库-技术培训-教材 IV.TP311.13

中国版本图书馆 CIP 数据核字(2004)第 083583 号

数据库与数据仓库

SHUJUKU YU SHUJU CANGKU

国家税务总局教材编写组 编

人民出版社 出版发行

(100706 北京朝阳门内大街 166 号)

河北省〇五印刷厂印刷 新华书店经销

2004 年 8 月第 1 版 2004 年 8 月北京第 1 次印刷

开本:787 毫米×960 毫米 1/16 印张:41.25

字数:687 千字 印数:1-5,000 册

ISBN 7-01-004489-9 定价:82.00 元

邮购地址 100706 北京朝阳门内大街 166 号

人民东方图书销售中心 电话 (010)65250042 65289539

序 言

由国家税务总局教育中心组织编写的全国税务系统基础知识、岗位分类和更新知识培训系列教材同大家见面了，我衷心地表示祝贺。这套系列培训教材的出版，为全国税务系统广泛深入地开展教育培训工作，全面提高广大税务干部的素质和能力，促进新时期税收事业的发展都将起到积极的作用。

党的十六大确定了本世纪头二十年我国全面建设小康社会的目标。要实现这一目标，需要各方面的共同努力，税收作为国民经济的重要杠杆之一，更要充分发挥好宏观调控作用。这对我们来说，既是机遇也是挑战。能否抓住机遇，迎接挑战，不负使命，关键在人，在于人的素质，而提高素质主要靠培训。为此，必须加大税务教育培训工作力度，按照全国组织工作会议要求，多层次、多渠道、大规模培训税务干部。通过税务教育培训工作，全面提高税务干部队伍的整体素质，圆满完成各项税收任务，为全面建设小康社会而努力奋斗。

教材是做好教育培训工作的基础，教材建设是教育培训工作中的重要组成部分。为全系统编写和提供高质量的教材，对于帮助广大税务干部提高自身素质和业务能力，加强队伍建设，都具有十分重要的意义。正是从这个意义上说，教材建设要锐意改革，勇于创新，与时俱进。要本着符合税收工作实际需要，符合税务教育培训与学习需要的原则，统一规划，认真组织，在求新、求变、求实上下功夫，多出精品佳作。

这套系列教材从策划、编审到出版，历时近三年，凝聚了税务系统400多名专家学者和业务骨干的心血。这套系列教材分为基础知识（X）、岗位专业知识与技能（Y）和更新知识（Z）三个部分，

形成 $X + Y + Z$ 的新型教材体系。总的看来，这套教材突破了传统教材的风格、模式和结构体系，实现了启发性与适用性、通俗性与趣味性的统一，组合灵活、简便适用，包含了全国税务系统公务员一般应具备的基本知识、各岗位所必须的专业技能以及新的知识和新的技能，也反映了税收工作的发展水平和改革方向。

希望广大税务干部加强学习，努力工作，不断提高理论素养、业务水平和工作能力，为新世纪的税收事业做出新的更大的贡献！

谈旭人

二〇〇三年四月二十一日

编 审 说 明

根据中共中央、国务院关于加强干部教育培训工作的要求和国家税务总局党组的指示，总局教育中心围绕建设一支政治过硬、业务熟练、作风优良的税务干部队伍的目标，注重培训教材建设，加强新形势下教材建设理论与实践的探索，确立了由基础知识（X）、岗位专业知识与技能（Y）、更新知识（Z）三个部分组成的X+Y+Z的新型教材体系。这套教材与税收工作紧密结合，通过大量典型案例和图表解释深奥的理论和复杂的问题，力求启发性与适用性、通俗性与趣味性相统一，是组织培训和干部自学的好帮手。

岗位专业知识与技能（Y）部分培训教材分为政策法规类、征收管理类、稽查类、计划会计统计类、信息管理类和综合类等类别。该部分教材针对税务工作各岗位应具备的专业知识与技能组织编写，突出实务性和可操作性，着重提高广大税务干部分析和解决实际问题的能力。

《数据库与数据仓库》为信息管理类教材。由顾伯群、王惠君负责具体策划指导，杨慧平、袁立炫、朱会彦、朱峻岭等参加编写，郭晓辉统稿，王秀、田绍槐主审。刘书明、夏日红、冷纪侔、刘建国、李芝麓、午锁平、单玉森、薛海波、钱志平、陈梦林、黎干等参加了教材的审定。

本书经国家税务总局教材编审委员会审定通过，同意出版发行。书中如有不妥之处，请读者批评指正。

国家税务总局教材编审委员会

二〇〇四年六月

《数据库与数据仓库》策划编审人员

总策划：许善达

策 划：程永昌 王 秀 陈小杭 王维平 孙 泽
顾伯群

协 助：高永清 杨国全

编 导：顾伯群 刘书明 夏日红 王惠君 郭晓辉

编 写：杨慧平 袁立炫 朱会彦 朱峻岭

统 稿：郭晓辉

主 审：王 秀 田绍槐

前 言

为了满足全国税务系统信息技术干部队伍教育培训工作的需要，切实提高信息技术岗位工作人员的业务素质，根据全国税务系统岗位专业知识与技能培训系列教材编写的要求，我们编写了 Y 系列信息管理类培训教材丛书。

丛书包括了《网络》、《硬件》、《中间件》、《工具软件》、《应用系统建设与维护》、《信息系统安全》以及《数据库与数据仓库》七本。编写过程中，我们通过会议、座谈等方式，认真听取了基层同志对丛书编写的意见和建议，并经过多次论证和反复修改，逐步形成了现有体系。Y 系列信息管理类培训教材丛书针对税务系统现有网络、硬件设备、操作系统、应用软件、信息安全以及数据库系统建设与维护的工作实际，充分把握“创新、务实、灵活”的原则，改变了传统信息技术教材的编写风格体系，在强调基本理论知识的同时，突出了实例分析、工作程序和工作方法的介绍，具有较强的实用性和可操作性。

丛书由国家税务总局教育中心顾伯群、王惠君，信息中心刘书明、夏日红、郭晓辉具体策划。《网络》分册由江苏省国税局刘建国、靳松、尚峻、葛以品、董文虎、潘正明、王晓培，甘肃省国税局午锁平、董立群、朱晓宁等同志编写；《硬件》分册由湖北省武汉市国税局陈锐、钱钢等同志编写；《中间件》分册由大连国税局冷纪侔、丁琳、曲直、刘福刚、董国承、张汇兵等同志编写；《工具软件》分册由河北省地税局岳轩、李同训等同志编写；《应用系统建设与维护》分册由国家税务总局信息中心杨慧平，深圳市国税局陈梦林、王晓明、翟小英、田仲昊、吴玉梅、蔡敬淳，深圳市地税局薛海波、彭文鸿等同志编写；《信息系统安全》分册由贵州省国税局李

芝麓、田炜，湖南省国税局黎干、张千等同志编写；《数据库与数据仓库》分册由国家税务总局信息中心杨慧平、袁立炫、朱会彦、郭晓辉和湖北省国税局朱峻岭等同志编写。

丛书由国家税务总局信息中心王秀主任主审，国家税务总局信息中心刘书明、夏日红、郭晓辉，扬州税务进修学院陆传基，湖南税务高等专科学校谢建全、田绍槐等同志参与审定。

丛书在编审过程中，得到国家税务总局教育中心和信息中心以及江苏、湖北、湖南、四川、贵州、甘肃省国税局，大连、深圳、武汉市国税局，河北省地税局，深圳市地税局，扬州税务进修学院，湖南税务高等专科学校领导及相关人员的大力支持和帮助，在此一并表示感谢。

由于编写工作量大、时间紧迫，加之全国税务系统信息化建设正处于不断改革和完善之中，本书中疏漏与不妥之处难免，敬请读者批评指正。我们的愿望，是努力打造出一套为税务系统信息技术人员“量身定做”的，切实满足信息技术知识学习需求的精品教材。

编 者

二〇〇四年六月

目 录

1 数据库基础	1
1.1 基本概念	1
1.2 关系数据库	16
1.3 关系数据库标准语言	28
1.4 关系系统及关系规范化	44
本章要点	57
思考与练习	57
2 数据仓库基础	64
2.1 数据仓库概述	64
2.2 数据仓库体系结构	75
2.3 数据仓库设计	82
2.4 联机分析处理 (OLAP)	100
2.5 数据挖掘	113
2.6 数据仓库系统工具	121
2.7 数据仓库系统工具的评价标准	182
本章要点	190
思考与练习	190
3 Oracle	192
3.1 PL/SQL	192
3.2 DBA	240
3.3 Oracle 性能优化与调整	322
3.4 Oracle 数据复制	333
本章要点	337
思考与练习	337
4 SYBASE	341
4.1 T-SQL	341

4.2 DBA	439
4.3 SYBASE 性能优化与调整	489
4.4 SYBASE 数据复制	515
本章要点	530
思考与练习	530
5 Microsoft SQL Server 7	535
5.1 Transact-SQL	535
5.2 SQL Server 数据库管理	583
5.3 SQL Server 性能优化与调整	627
5.4 SQL Server 数据复制	635
本章要点	646
思考与练习	647

1 数据库基础

本章比较详细地讲述了数据库的基本概念、关系数据库理论和 SQL 语言基础, 为学习 Oracle、Sybase、MS SQL Server 等商业数据库系统提供了理论基础。

1.1 基本概念

1.1.1 引言

1. 数据、数据库、数据库系统、数据库管理系统

● 数据 (Data)

数据实际上就是描述事物的符号记录。

文字、图形、图像、声音、学生的档案记录、货物的运输情况……, 这些都是数据。数据的形式本身并不能完全表达其内容, 需要经过语义解释。数据与其语义是不可分的。

● 数据库 (Database, DB)

数据库是存储在计算机内有结构的大量的共享的数据集合。

它可以供各种用户共享、具有最小冗余度和较高的数据独立性。DBMS 在数据库建立、运用和维护时对数据库进行统一控制, 以保证数据的完整性、安全性, 并在多用户同时使用数据库时进行并发控制, 在发生故障后对系统进行恢复。

● 数据库管理系统 (Database Management System, DBMS)

数据库管理系统是位于用户与操作系统之间的一层数据管理软件。

数据库在建立、运用和维护时由数据库管理系统统一管理、统一控制。数据库管理系统使用户能方便地定义和操纵数据, 并能够保证数据的安全性、完整性、多用户对数据的并发使用及发生故障后的系统恢复。

● 数据库系统 (Database System, DBS)

数据库系统是指在计算机系统中引入数据库后的系统构成，一般由数据库、数据库管理系统（及其开发工具）、应用系统、数据库管理员和用户构成。

● 数据库系统管理员（Database Administrator, DBA）

数据库系统管理员负责数据库的建立、使用和维护的专门人员。

2. 数据库技术的产生与发展

数据库的发展阶段如表 1-1 所示。

表 1-1 数据管理三个阶段的比较

		人工管理	文件系统	数据库系统
背景	应用背景	科学计算	科学计算、管理	大规模管理
	硬件背景	无直接存取存储设备	磁盘、磁鼓	大容量磁盘
	软件背景	没有操作系统	有文件系统	有数据库管理系统
	处理方式	批处理	联机实时处理批处理	联机实时处理，分布处理批处理
特点	数据的管理者	人	文件系统	数据库管理系统
	数据面向的对象	某一应用程序	某一应用程序	整个应用系统
	数据的共享程度	无共享，冗余度极大	共享性差，冗余度大	共享性高，冗余度小
	数据的独立性	不独立，完全依赖于程序	独立性差	具有高度的物理独立性和逻辑独立性
	数据的结构化	无结构	记录内有结构，整体无结构	整体结构化，用数据模型描述
	数据控制能力	应用程序自己控制	应用程序自己控制	由数据库管理系统提供数据安全性、完整性、并发控制和恢复能力

以上表格中用到的术语：

● 数据结构化

数据结构化是数据库与文件系统的根本区别。在描述数据时不仅要描述数据本身，还要描述数据之间的联系。

- 数据的共享性

数据库系统从整体的角度来看待和描述数据，数据不再面向某个应用而是面向整个系统。

- 数据冗余度

指同一数据重复存储时的重复程度。

- 数据的一致性

指同一数据不同拷贝时的值一样（采用人工管理或文件系统管理时，由于数据被重复存储，当不同的应用使用和修改不同的拷贝时就易造成数据的不一致）。

- 物理独立性

当数据的存储结构（或物理结构）改变时，通过对映象的相应改变可以保持数据的逻辑结构可以不变，从而应用程序也不必改变。

- 逻辑独立性

当数据的总体逻辑结构改变时，通过对映象的相应改变可以保持数据的局部逻辑结构不变，应用程序是依据数据的局部逻辑结构编写的，所以应用程序不必修改。

- 数据的安全性（Security）

数据的安全性是指保护数据，防止不合法使用数据造成数据的泄密和破坏，使每个用户只能按规定，对某些数据以某些方式进行访问或处理。

- 数据的完整性（Integrity）

数据的完整性指数据的正确性、有效性和相容性。即将数据控制在有效的范围内，或要求数据之间满足一定的关系。

- 并发（Concurrency）控制

当多个用户的并发进程同时存取、修改数据库时，可能会发生相互干扰而得到错误的结果并使得数据库的完整性遭到破坏，因此必须对多用户的并发操作加以控制和协调。

- 数据库恢复（Recovery）

计算机系统的硬件故障、软件故障、操作员的失误以及故意的破坏也会影响数据库中数据的正确性，甚至造成数据库部分或全部数据的丢失。DBMS 必

须具有将数据库从错误状态恢复到某一已知的正确状态（亦称为完整状态或一致状态）的功能。

3. 数据库技术的研究领域

● 数据库管理系统软件的研制

DBMS 的研制包括 DBMS 本身的研制及以 DBMS 为核心的一组相互联系的软件系统。研制的目标是扩大功能、提高性能和提高用户的生产率。

● 数据库设计

数据库设计的主要任务是在 DBMS 的支持下，按照应用的要求，为某一部门或组织设计一个结构合理、使用方便、效率较高的数据库及其应用系统。

● 数据库理论

数据库理论的研究主要集中于关系的规范化理论、关系数据理论等。近年来，随着人工智能与数据库理论的结合、并行计算机的发展，数据库逻辑演绎和知识推理、并行算法等理论研究，以及演绎数据库系统、知识库系统的研制都已成为新的研究方向。

1.1.2 数据模型

在数据库中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。通俗地讲数据模型就是现实世界的模拟。

数据模型分为两个不同的层次：

● 概念模型

也称信息模型，它是按用户的观点对数据和信息进行建模。

● 数据模型

主要包括网状模型、层次模型、关系模型等，它是按计算机系统的观点对数据进行建模。

1. 概念模型

概念模型用于信息世界的建模。概念模型不依赖于某一个 DBMS 支持的数据模型。概念模型可以转换为计算机上某一 DBMS 支持的特定数据模型。

概念模型特点：

(1) 具有较强的语义表达能力，能够方便、直接地表达应用中的各种语义知识。

(2) 应该简单、清晰、易于用户理解，是用户与数据库设计人员之间进

行交流的语言。

信息世界中的基本概念主要包含以下几个部分。

- 实体 (Entity)

客观存在并可相互区别的事物称为实体。实体可以是具体的人、事、物，也可以是抽象的概念或联系。

- 属性 (Attribute)

实体所具有的某一特性称为属性。一个实体可以由若干个属性来描述。

- 码 (Key)

惟一标识实体的属性集称为码。

- 域 (Domain)

属性的取值范围称为该属性的域。

- 实体型 (Entity Type)

用实体名及其属性名集合来抽象和描述同类实体，称为实体型。

- 实体集 (Entity Set)

同型实体的集合称为实体集。

- 联系 (Relationship)

现实世界中事物内部以及事物之间的联系，在信息世界中反映为实体内部的联系和实体之间的联系。

- ◆ 一对一联系 (1: 1)

如果对于实体集 A 中的每一个实体，实体集 B 中至多有一个实体与之联系，反之亦然，则称实体集 A 与实体集 B 具有一对一联系，记为 1: 1。

- ◆ 一对多联系 (1: n)

如果对于实体集 A 中的每一个实体，实体集 B 中有 n 个实体 ($n \geq 0$) 与之联系，反之，对于实体集 B 中的每一个实体，实体集 A 中至多只有一个实体与之联系，则称实体集 A 与实体 B 有一对多联系，记为 1: n 。

- ◆ 多对多联系 ($m: n$)

如果对于实体集 A 中的每一个实体，实体集 B 中有 n 个实体 ($n \geq 0$) 与之联系，反之，对于实体集 B 中的每一个实体，实体集 A 中也有 m 个实体 ($m \geq 0$) 与之联系，则称实体集 A 与实体 B 具有多对多联系，记为 $m: n$ 。

实体型之间的一对一、一对多、多对多联系不仅存在于两个实体型之间，也存在于两个以上的实体型之间。