

Access 2016 数据库应用技术 案例教程学习指导

主 编 林敦欣 刘 垣
副主编 郭李华 徐沛然 王一蕾

清华大学出版社
北 京

内 容 简 介

本书是《Access 2016 数据库应用技术案例教程》(ISBN: 978-7-302-69315-4)的配套学习指导,其中,第1章、第3章至第9章以高校“教务管理”数据库为基础,分别介绍数据库系统基础知识、Access 2016 数据库和表、查询、窗体、报表、宏、VBA 程序设计、ADO 数据库编程等内容;第2章介绍“人工智能+”数据技术,以国产大模型赋能课程学习为实验案例;第10章为数据库应用系统开发案例;附录以二维码形式给出常用的操作命令、内部函数等内容。

本书具有双重教学功能定位:既可作为人工智能时代高等院校数据库应用技术类课程的学习指导用书,帮助学生系统掌握数据库理论与实践技能;也适用于全国计算机二级 Access 数据库程序设计考试的备考,并可作为培训教材或实验参考用书,助力考生高效复习、精准把握考试要点,切实提升应试能力与数据库技术应用素养。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。举报:010-62782989, beiqinquan@tup.tsinghua.edu.cn。

图书在版编目(CIP)数据

Access 2016 数据库应用技术案例教程学习指导 / 林敦欣, 刘垣

主编.-- 北京:清华大学出版社,2025.6.

ISBN 978-7-302-69316-1

I. TP311.132.3

中国国家版本馆 CIP 数据核字第 2025KY3499 号

责任编辑:王 定

封面设计:周晓亮

版式设计:思创景点

责任校对:成凤进

责任印制:杨 艳

出版发行:清华大学出版社

网 址: <https://www.tup.com.cn>, <https://www.wqxuetang.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社 总 机:010-83470000 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课 件 下 载: <http://www.tup.com.cn>, 010-62794504

印 装 者:北京瑞禾彩色印刷有限公司

经 销:全国新华书店

开 本:185mm×260mm 印 张:12 字 数:386千字

版 次:2025年7月第1版 印 次:2025年7月第1次印刷

定 价:59.80元

产品编号:110644-01

前言

P R E F A C E

革故鼎新，笃行不怠。在课程素养教育理念指引下，我们围绕工程教育认证标准，探索了“人工智能+”与数据库技术的结合，并据此重新设置了原版教材的内容架构。

1992年11月，微软公司首次推出 Access。此后，Access 历经 14 个版本的变迁，最新版本为 Access 2024。自 Access 2007 起，其数据库文件格式始终未变。微软公司也明确表示，在后续的 Access 版本升级中，暂无更改数据库文件格式的计划。本书是《Access 2016 数据库应用技术案例教程》(ISBN: 978-7-302-69315-4)的配套学习指导，由 10 章正文和 4 个附录构成。第 1 章、第 3 章至第 9 章分别介绍了数据库系统基础知识、Access 2016 数据库和表、查询、窗体、报表、宏、VBA 程序设计与 ADO 数据库编程，各章均以大学教务管理数据库为基础。第 2 章介绍了“人工智能+”数据技术，实验案例是国产大模型赋能课程学习。第 10 章为数据库应用系统开发案例，实验案例的前 5 个是综合操作训练，后 4 个是数据库应用程序的开发。本书各附录内容丰富且实用：附录 A 呈现教务管理数据库各表的结构与记录；附录 B 收录常用字符与 ASCII 码对照表；附录 C 整理常用宏操作命令；附录 D 梳理 VBA 常用内部函数。

本书每章的末尾附有多个实验案例，这些实验案例既有基础验证型和综合设计型案例，也有需要调研、团队合作才能完成的创新研究型案例。读者通过这些案例的学习，举一反三加以迁移，就能解决实际生活和工作中的许多问题。

本书为福建省“十四五”普通高等教育本科规划教材建设项目的成果。编者都是高校计算机类教学一线教师，实验案例中的许多内容都是他们多年教学实践的总结。本书由林敦欣、刘垣任主编，郭李华、徐沛然、王一蕾任副主编，参与编写工作的还有苏备迎、温馨、韩宜航、郑兆铨和陈治杰等人，本书的完成也离不开原版作者连贻捷、刘琰、林铭德和张波尔等老师的贡献。本书的编写得到了福建理工大学、福州大学、湖北工程学院、福建农林大学等多所院校的大力支持，在此一并表示感谢！

由于编者水平有限，书中难免存在疏漏和不足，敬请读者提出宝贵意见和建议。

本书提供实验案例源文件、思考与练习参考答案、模拟试卷(模拟全国计算机等级考试二级 Access 数据库程序设计考题格式)、模拟试卷参考答案，读者可扫下列二维码学习。



实验案例
源文件



思考与练习
参考答案



模拟试卷



模拟试卷
参考答案

编 者

2025年3月于榕城旗山

目 录

C O N T E N T S

第 1 章 数据库系统概述 ····· 1	第 2 章 “人工智能+” 数据技术 ····· 20
1.1 知识要点····· 1	2.1 知识要点····· 20
1.1.1 数据管理技术的产生与发展····· 1	2.1.1 人工智能的定义····· 20
1.1.2 数据库技术的基本术语····· 2	2.1.2 人工智能的起源与发展····· 21
1.1.3 数据库系统的三级模式结构····· 3	2.1.3 人工智能的主要学派····· 22
1.1.4 国产数据库····· 4	2.1.4 机器学习····· 23
1.1.5 由现实世界到数据世界····· 4	2.1.5 深度学习····· 24
1.1.6 数据模型的分层····· 5	2.1.6 AI 的技术发展方向····· 24
1.1.7 概念模型和 E-R 图····· 5	2.1.7 我国 AI 产业发展现状····· 25
1.1.8 关系模型····· 6	2.1.8 我国 AI 标准化体系建设总体 要求····· 25
1.1.9 关系运算····· 6	2.1.9 我国人工智能标准体系结构····· 25
1.1.10 关系的完整性····· 7	2.1.10 我国人工智能治理····· 25
1.1.11 数据库设计····· 7	2.1.11 国家数据基础设施内涵····· 26
1.2 思考与练习····· 8	2.1.12 国家数据基础设施发展愿景 目标····· 26
1.2.1 选择题····· 8	2.1.13 国家数据基础设施总体技术 架构····· 26
1.2.2 填空题····· 11	2.1.14 我国数据库产业发展现状····· 26
1.2.3 简答题····· 12	2.1.15 我国数据库支撑体系····· 27
1.3 实验案例····· 14	2.1.16 我国数据库关键技术发展 趋势····· 27
实验案例 1····· 14	2.2 思考与练习····· 27
实验案例 2····· 14	2.2.1 选择题····· 27
实验案例 3····· 15	2.2.2 填空题····· 28
实验案例 4····· 15	2.2.3 简答题····· 28
实验案例 5····· 15	2.3 实验案例····· 28
实验案例 6····· 16	实验案例 1····· 28
实验案例 7····· 16	实验案例 2····· 30
实验案例 8····· 17	
实验案例 9····· 17	
实验案例 10····· 18	
实验案例 11····· 18	

实验案例 3	31	4.2.1 选择题	61
实验案例 4	32	4.2.2 填空题	64
实验案例 5	34	4.2.3 简答题	64
第 3 章 数据库和表	36	4.3 实验案例	65
3.1 知识要点	36	实验案例 1	65
3.1.1 创建数据库	36	实验案例 2	66
3.1.2 数据库的基本操作	36	实验案例 3	66
3.1.3 表的视图	37	实验案例 4	67
3.1.4 创建表	37	实验案例 5	68
3.1.5 导入数据与链接数据	38	实验案例 6	69
3.1.6 主键的设置	38	实验案例 7	70
3.1.7 数据类型	38	实验案例 8	71
3.1.8 字段属性	38	实验案例 9	71
3.1.9 建立表之间的关系	40	实验案例 10	72
3.1.10 编辑数据表	41	实验案例 11	73
3.2 思考与练习	42	实验案例 12	74
3.2.1 选择题	42	实验案例 13	74
3.2.2 填空题	45	实验案例 14	75
3.2.3 简答题	46	实验案例 15	76
3.3 实验案例	46	实验案例 16	76
实验案例 1	46	第 5 章 窗体	78
实验案例 2	47	5.1 知识要点	78
实验案例 3	51	5.1.1 窗体概述	78
实验案例 4	53	5.1.2 创建窗体	79
实验案例 5	55	5.1.3 设计窗体	79
第 4 章 查询	57	5.1.4 修饰窗体	83
4.1 知识要点	57	5.1.5 定制用户入口界面	84
4.1.1 查询的功能	57	5.2 思考与练习	84
4.1.2 查询的类型	57	5.2.1 选择题	84
4.1.3 查询视图	58	5.2.2 简答题	86
4.1.4 使用向导创建查询	59	5.3 实验案例	86
4.1.5 条件表达式	59	实验案例 1	86
4.1.6 汇总计算	60	实验案例 2	88
4.1.7 SQL 查询	60	实验案例 3	89
4.2 思考与练习	61	实验案例 4	91
		实验案例 5	92

实验案例 6	93	第 8 章 VBA 程序设计	122
实验案例 7	93	8.1 知识要点	122
实验案例 8	94	8.1.1 程序设计语言	122
第 6 章 报表	96	8.1.2 VBA 概述	123
6.1 知识要点	96	8.1.3 数据类型、表达式和函数	124
6.1.1 报表的作用及类型	96	8.1.4 程序控制结构	130
6.1.2 快速创建报表的方法	97	8.1.5 过程与函数	133
6.1.3 使用设计视图创建和编辑 报表	98	8.2 思考与练习	134
6.1.4 创建图表报表和标签报表	101	8.2.1 选择题	134
6.1.5 报表的导出	102	8.2.2 填空题	139
6.2 思考与练习	102	8.2.3 简答题	141
6.2.1 选择题	102	8.2.4 程序设计题	141
6.2.2 简答题	104	8.3 实验案例	142
6.3 实验案例	104	实验案例 1	142
实验案例 1	104	实验案例 2	143
实验案例 2	105	实验案例 3	144
实验案例 3	106	实验案例 4	145
实验案例 4	108	实验案例 5	146
实验案例 5	110	实验案例 6	148
第 7 章 宏	113	实验案例 7	149
7.1 知识要点	113	实验案例 8	150
7.1.1 宏的作用及类型	113	实验案例 9	151
7.1.2 宏的设计与运行	114	实验案例 10	153
7.1.3 使用宏创建菜单	114	第 9 章 ADO 数据库编程	155
7.2 思考与练习	115	9.1 知识要点	155
7.2.1 选择题	115	9.1.1 数据库引擎和接口	155
7.2.2 简答题	116	9.1.2 ADO	156
7.3 实验案例	117	9.1.3 ADO 主要对象	156
实验案例 1	117	9.1.4 ADO 在 Access 中的应用	159
实验案例 2	118	9.2 思考与练习	160
实验案例 3	119	9.2.1 选择题	160
实验案例 4	120	9.2.2 填空题	161
实验案例 5	121	9.2.3 简答题	161
		9.2.4 程序设计题	161
		9.3 实验案例	162
		实验案例 1	162

实验案例 2	164	10.2.2 填空题	171
实验案例 3	165	10.2.3 简答题	171
实验案例 4	165	10.3 实验案例	171
第 10 章 数据库应用系统开发	168	实验案例 1	171
10.1 知识要点	168	实验案例 2	172
10.1.1 软件	168	实验案例 3	173
10.1.2 软件测试	168	实验案例 4	173
10.1.3 程序调试	169	实验案例 5	174
10.1.4 什么是大模型	169	实验案例 6	175
10.1.5 大模型的发展	169	实验案例 7	177
10.1.6 大模型应用体验	169	实验案例 8	178
10.1.7 大模型辅助开发案例	169	实验案例 9	179
10.2 思考与练习	170	参考文献	181
10.2.1 选择题	170	附录	182

数据库系统概述

1.1 知识要点

1.1.1 数据管理技术的产生与发展

数据管理是指对数据进行分类、组织、编码、存储、检索、维护和应用，它是数据处理的中心问题。随着应用需求的推动和计算机软硬件的发展，数据管理技术经历了人工管理、文件系统和数据库系统三个阶段(后又发展为分布式数据库系统和面向对象数据库系统等)。

1. 人工管理阶段

此阶段主要是指 20 世纪 50 年代中期以前，数据需要由应用程序定义和管理，一个数据集只能对应一个应用程序。数据无共享，冗余度极大；数据不独立，完全依赖于程序。这个阶段的计算机很简陋，主要应用于科学计算。

2. 文件系统阶段

此阶段主要是指 20 世纪 50 年代末到 60 年代中期。在这一阶段，人们利用专门的数据管理软件(即文件系统)管理数据。对于一个特定的应用，数据被集中组织存放在多个数据文件或文件组中。人们为了更好地管理和利用这些数据，会针对该文件组开发特定的应用程序。然而，这种数据管理方式存在问题：数据的共享性差，冗余度大；数据独立性差。此时计算机除了应用于科学计算，也开始应用于数据管理。

3. 数据库系统阶段

自 20 世纪 60 年代末期以来都属于此阶段。有专门的数据管理软件(DBMS)对数据库

提供安全性、完整性、并发控制等支持。数据共享性高，冗余度小；数据具有高度的物理独立性和一定的逻辑独立性；数据整体结构化，用数据模型描述。

伴随着应用需求的推动和计算机软硬件技术的发展，数据库系统阶段出现了多种数据库：关系数据库、并行数据库、分布式数据库、对象-关系数据库、面向对象数据库、以互联网大数据应用为背景发展起来的分布式非关系型的数据库管理系统(NoSQL)等。分布式数据库系统由数据库技术与网络通信技术相结合而产生，面向对象的数据库系统由数据库技术与面向对象程序设计技术相结合而产生。

4. 中国数据管理的发展

1956年，周恩来总理领导制定《1956—1967年科学技术发展远景规划纲要》，同年8月我国成立中国科学院计算技术研究所筹委会，由数学家华罗庚任主任委员。

1978年，萨师煊先生首次在中国人民大学开设数据库课程。1999年8月中国计算机学会数据库专业委员会成立，标志着中国数据库领域进入了一个组织化、专业化发展的新阶段。

2021年6月，中国电子信息行业联合会发布《数据管理从业人员能力等级要求》团体标准及编制说明的公告，将数据管理从业人员分为首席数据官、资深数据管理工程师、数据管理工程师、助理数据管理工程师四个等级；第十三届全国人民代表大会常务委员会第二十九次会议通过《中华人民共和国数据安全法》，该法旨在规范数据处理活动，保障数据安全，促进数据开发利用，保护个人、组织的合法权益，维护国家主权、安全和发展利益。

2022年12月，中共中央、国务院发布《关于构建数据基础制度更好发挥数据要素作用的意见》，以数据产权、流通交易、收益分配、安全治理为重点，初步提出我国数据基础制度20条政策举措，简称“数据二十条”。它是我国数据基础制度的“四梁八柱”。

2023年3月，中共中央、国务院印发《党和国家机构改革方案》，组建由国家发展和改革委员会管理的国家数据局，负责协调推进数据基础制度建设，统筹数据资源整合共享和开发利用，统筹推进数字中国、数字经济、数字社会规划和建设等。

2024年1月，由国家数据局等17部门联合发布《“数据要素×”三年行动计划(2024—2026年)》。2024年8月，国务院第40次常务会议通过《网络数据安全条例》，自2025年1月1日起施行。

1.1.2 数据库技术的基本术语

1. 数据(data)

数据是数据库中存储的基本对象，是描述事物的符号记录。数据通常分为数值型数据和非数值型数据两种形式。每个数据都有其语义。用表格描述的数据称为结构化数据。

数据是国家基础性战略资源和重要生产要素，是数字化、网络化、智能化的基础，已快速融入生产、分配、流通、消费和社会服务管理等各环节，深刻改变着生产方式、生活方式和社会治理方式。

信息以数据为载体，是具有一定含义的、经过加工处理的数据，是客观事物存在方式和运动状态的反映，对人类决策有帮助和价值。例如，气象台依据事先勘测采集的气压、

云层、温度、湿度、风力等数据，经过整理加工和综合分析得出的天气预报即为信息。

2. 大数据(big data)

依据 GB/T 35295-2017，大数据指具有体量巨大、来源多样、生成极快、多变等特征并且难以用传统数据体系结构有效处理的包含大量数据集的数据。2014年3月，“大数据”首次被写入《政府工作报告》。

3. 数据库(database, DB)

数据库是长期存储在计算机内，有组织且可共享的大量数据的集合。数据库中不仅存放数据，还存放数据与数据之间的联系。

随着数据库的发展，出现了数据仓库，数据仓库是一个面向主题、集成、非易失性和随时间变化的集合，用于支持管理层的决策。

4. 数据库管理系统(database management system, DBMS)

DBMS 是数据库系统的核心组成部分，是位于用户与操作系统之间的一层数据管理软件，用于描述、管理和维护数据库。

DBMS 的主要功能：数据定义、组织、存储和管理功能；数据操纵功能；数据库的建立和维护功能。当今主流的数据库管理系统是关系数据库管理系统(RDBMS)。

5. 数据库系统(database system, DBS)

DBS 是由数据库、数据库管理系统、数据库应用系统和用户组成的存储、管理、处理和维持数据的系统，其中用户又分为终端用户、应用程序员、系统分析员、数据库设计人员和数据库管理员 DBA 等多种。

数据库管理员负责全面管理和控制数据库系统，其主要工作是：数据库设计、数据库维护、改善系统性能，提高系统效率。

DBS 的数据有安全性保护和完整性检查措施，能并发控制和恢复数据库，数据具有共享性高、冗余度低，独立性高的特点。数据独立性一般分为逻辑独立性和物理独立性两种。逻辑独立性是指用户的应用程序与数据库的逻辑结构相互独立；物理独立性是指用户的应用程序与数据库中数据的物理存储相互独立。

1.1.3 数据库系统的三级模式结构

从数据库应用开发者角度看，数据库系统通常采用“外模式-模式-内模式”三级模式结构，相邻两级结构之间的两层映像是外模式/模式映像、模式/内模式映像。这两层映像保证了数据库系统中的数据能够具有较高的逻辑独立性和物理独立性。

数据库系统的三级模式是对数据的三个抽象级别，它把数据的具体组织留给 DBMS 管理，使用户能逻辑地、抽象地处理数据，从而实现了数据的独立性，即当数据的结构和存储方式发生变化时，应用程序不受影响。

1. 外模式(external schema)

外模式又称用户模式或子模式，是数据库用户能够看见并使用的局部数据的逻辑结构和特征的描述，是与某一应用有关的数据的逻辑表示。

外模式是各个用户的数据视图，如果不同的用户在不同应用需求、看待数据的方式、对数据保密的要求等方面存在差异，则其外模式的描述就不同。一个数据库可以有多个外模式。

2. 模式(schema)

模式又称逻辑模式或概念模式，是数据库中全体数据的逻辑结构和特征的描述，是所有用户的公共数据视图。一个数据库只有一个模式。

3. 内模式(internal schema)

内模式又称存储模式，是数据物理结构和存储方式的描述，是数据在数据库内部的组织方式。一个数据库只有一个内模式。

1.1.4 国产数据库

1. 国产数据库的发展历程

起步阶段(1978—2000年)、跟踪阶段(2000—2008年)、追赶阶段(2008—2014年)、并跑阶段(2014年至今)。

2. 两个国产数据库

蚂蚁科技集团股份有限公司的分布式关系数据库 OceanBase，华为技术有限公司的 openGauss 数据库。

1.1.5 由现实世界到数据世界

获得一个 DBMS 所支持的数据模型的过程，是一个从现实世界的事物出发，经过人们的抽象，以获得人们所需要的概念模型和数据模型的过程。信息在这个过程中经历了三个不同的世界：现实世界、概念世界和数据世界。

1. 现实世界

现实世界是人们通常所指的客观世界，事物及其联系就处在这个世界中。一个实际存在并且可以识别的事物称为个体，个体可以是一个具体的事物，如一个学生、一所学校，也可以是一个抽象的概念，如某位学生的特长与爱好。通常把具有相同特征个体的集合称为全体。

2. 概念世界

概念世界又称为信息世界，是指现实世界的客观事物经人们综合分析后，在头脑中形成的印象与概念。现实世界中的个体和全体在概念世界中分别称为实体和实体集。概念世界不是现实世界在人脑的简单主观反映，而是经过选择、命名、分类等抽象过程产生的概念模型。

3. 数据世界

数据世界又称为机器世界或计算机世界。进入计算机的信息必须是数字化的。当信息由概念世界进入数据世界后，概念世界的实体和属性等在数据世界中要进行数字化的表示，每个实体和实体集在数据世界中分别称为记录和文件。

1.1.6 数据模型的分层

数据库的类型是依据数据模型来划分的，数据模型是数据库系统的基础。数据模型由数据结构、数据操作与数据的约束条件 3 部分组成。根据数据抽象的不同级别，可以将数据模型分为：概念数据模型、逻辑数据模型和物理数据模型。

1. 概念数据模型(Conceptual Data Model, CDM)

概念数据模型简称为概念模型或信息模型，是按用户的观点或认识对现实世界的数据和信息进行建模，主要用于数据库设计。常用的概念模型表示方法是 E-R 图。

2. 逻辑数据模型(Logical Data Model, LDM)

逻辑层是数据抽象的第二层抽象，用于描述数据库数据的整体逻辑结构。该层的数据抽象称为逻辑数据模型，简称为逻辑模型，也可以称为数据模型。

不同的 DBMS 提供不同的逻辑数据模型，例如，层次模型、网状模型、关系模型、面向对象模型、对象关系模型、半结构化模型等，其中层次模型和网状模型统称为格式化模型。

- **层次模型**：最早出现的数据模型，用树状结构来表示各类实体及实体之间的联系。
- **网状模型**：用网状结构来表示各类实体及实体之间的联系。
- **关系模型**：用规范化的二维表来表示各类实体及实体之间的联系。
- **半结构化模型**：随着互联网的迅速发展，Web 上各种半结构化、非结构化数据源已成为重要的信息来源，产生了以 XML 为代表的半结构化数据模型和非结构化数据模型。

3. 物理数据模型(Physical Data Model, PDM)

物理层的数据抽象称为物理数据模型，简称为物理模型，它不但由 DBMS 的设计决定，而且与操作系统、计算机硬件密切相关。

1.1.7 概念模型和 E-R 图

概念模型是对信息世界建模，是现实世界到概念世界的第一层抽象。最常用的概念模型表示方法是实体-联系方法，又称 E-R 图或 E-R 方法或 E-R 模型。E-R 图是一种语义模型，是现实世界到信息世界的事物及事物之间关系的抽象表示。

E-R 图是不受任何 DBMS 约束的面向用户的表达方法，能够直观表示现实世界中的客观实体、属性及实体之间的联系。构成 E-R 图的基本要素是实体型、属性和联系。相关术语如下。

- **实体**：客观世界中可区别于其他事物的“事物”或“对象”。
- **实体集**：指具有相同类型及相同性质或属性的实体集合。
- **实体型**：用实体名及其属性名集合来抽象和刻画同类实体。
- **属性**：是实体集中每个实体都具有的特征描述。
- **码**：又称键，能唯一标识实体的属性或属性集。
- **域**：一个属性所允许的取值范围或集合称为该属性的域。

- **实体之间的联系：**实体之间的对应关系称为联系，它反映了现实世界事物之间相互关联的情况。联系分为：一对一联系(1:1)、一对多联系(1:n)和多对多联系(m:n)。

1.1.8 关系模型

关系数据库系统采用关系模型作为数据的组织方式。关系模型于 1970 年由埃德加·弗兰克·科德首次提出，它是一种用二维表表示实体集，主码标识实体，外码表示实体间联系的数据模型。

1. 基本术语

- **关系：**对应通常所说的二维表，它由行和列组成，且必须满足一定的规范条件。
- **关系名：**每个关系的名称。
- **元组：**二维表中的每一行称为关系的一个元组，它对应于实体集中的一个实体。
- **属性：**二维表中的每一列对应于实体的一个属性，每个属性要有一个属性名。
- **值域：**每个属性的取值范围。
- **分量：**元组中的一个属性值。
- **候选码：**若关系中的某一属性组的值能唯一标识一个元组，则称该属性组为候选码。
- **主码：**也称主键或关键字。如果一个关系有多个候选码，则选定其中一个为主码。
- **外码：**也称外键或外部关键字。为了实现表与表之间的联系，通常将一个表的主码作为数据之间联系的纽带放到另一个表中，这个起联系作用的属性称为外码。
- **关系模式：**对关系的描述，一般表示为：关系名(属性 1, 属性 2, …, 属性 n)

2. 关系模型的性质

关系是建立在严格数学理论基础之上的二维表。一张二维表中的元组和属性的个数都是有限的，且与次序无关；元组具有唯一性，属性名也是唯一的；元组分量具有原子性，分量的值取自同一个域。

1.1.9 关系运算

关系运算是针对关系数据库的数据操纵。关系模型中常用的关系操作包括查询、插入、删除、修改。查询是关系操作中最主要的部分。查询操作可分为并、差、交、广义笛卡尔积、选择、投影、连接、除等。

关系代数用对关系的运算来表达查询。关系代数的运算对象是关系，运算结果也是关系。根据运算符的不同，关系代数的运算分为传统的集合运算和专门的关系运算。

1. 传统的集合运算

设关系 R 和关系 S 具有相同的目 n(即两个关系都有 n 个属性)，且相应的属性取自同一个域，t 是元组变量， $t \in R$ 表示 t 是 R 的一个元组。

- **并：**关系 R 和关系 S 的并记作： $R \cup S = \{t | t \in R \vee t \in S\}$ 。其结果仍为 n 目关系，由属于 R 或属于 S 的元组组成。
- **差：**关系 R 和关系 S 的差记作： $R - S = \{t | t \in R \wedge t \notin S\}$ 。其结果仍为 n 目关系，由属

于 R 但不属于 S 的所有元组组成。

- **交**：关系 R 和关系 S 的交记作： $R \cap S = \{t \mid t \in R \wedge t \in S\}$ 。其结果仍为 n 目关系，由既属于 R 又属于 S 的元组组成。关系的交可以用差来表示，即 $R \cap S = R - (R - S)$ 。
- **广义笛卡尔积**：两个分别为 n 目和 m 目的关系 R 和关系 S 的笛卡尔积是一个 (n+m) 列的元组的集合。元组的前 n 列是关系 R 的一个元组，后 m 列是关系 S 的一个元组。若 R 有 k_1 个元组，S 有 k_2 个元组，则关系 R 和关系 S 的笛卡尔积有 $k_1 \times k_2$ 个元组。记作： $R \times S = \{t_r t_s \mid t_r \in R \wedge t_s \in S\}$ 。

2. 专门的关系运算

- **选择**：根据给定的条件，从一个关系中选出一个或多个元组，即二维表中的行。
- **投影**：从一个关系中选择某些特定的属性(表中的列)，重新排列后组成一个新的关系。
- **连接**：从两个或多个关系中选取属性间满足一定条件的元组，组成一个新的关系。

1.1.10 关系的完整性

关系模型的完整性规则是为保证数据库中数据的正确性和相容性，对关系模型提出的某种约束条件或规则。完整性通常包括实体完整性、参照完整性和用户自定义完整性，其中实体完整性和参照完整性是关系模型必须满足的完整性约束条件。

1. 实体完整性(entity integrity)

实体完整性是指关系的主码不能重复，也不能取空值 null。

2. 参照完整性(referential integrity)

参照完整性是定义建立关系之间联系的主码与外码引用的约束条件。

3. 用户自定义完整性(user-defined integrity)

用户自定义完整性是针对不同应用领域的语义，由用户自己定义的一些完整性约束条件。

1.1.11 数据库设计

数据库设计是数据库及其应用系统的设计，它是一项软件工程，开发过程遵循软件工程的一般原理和方法。数据库设计目前一般采用生命周期法，将整个数据库应用系统的开发分解成目标独立的 6 个阶段：需求分析阶段、概念结构设计阶段、逻辑结构设计阶段、物理结构设计阶段、数据库实施阶段，以及数据库运行和维护阶段。设计一个完善的数据库应用系统往往是这 6 个阶段的不断反复。

1. 需求分析

需求分析是整个数据库设计过程的基础，需要与用户有效交流，这是最困难和最耗时的一步。需求分析的结果是否准确反映用户的实际要求，将直接影响数据库应用系统的质量。

2. 概念结构设计

概念结构设计阶段是将需求分析得到的用户需求抽象为概念模型的过程。E-R 图是此阶段数据库设计中广泛使用的数据库建模工具。

3. 逻辑结构设计

逻辑结构设计的任务是把概念结构设计阶段得到的 E-R 图转换为逻辑结构, 这个逻辑结构要与选用的 DBMS 产品的数据模型相符合。当前的数据库应用系统大都采用支持关系数据模型的 RDBMS。

E-R 图向关系模型转换遵循的原则: 一个实体型转换为一个关系模式; 一个 1:1 联系可以转换为一个独立的关系模式, 也可以与任意一端对应的关系模式合并; 一个 1:n 联系可以转换为一个独立的关系模式, 也可以与 n 端对应的关系模式合并; 一个 m:n 联系可以转换为一个关系模式, 与该联系相连的各实体的码及联系本身的属性均转换为关系的属性; 三个或三个以上实体间的一个多元联系可以转换为一个关系模式; 具有相同码的关系模式可以合并。

4. 物理结构设计

数据库在物理设备上的存储结构与存取方法称为数据库的物理结构, 它依赖于选定的 DBMS。数据库的物理结构设计是为一个给定的逻辑数据模型选取一个最适合应用要求的物理结构的过程。

5. 数据库实施

完成物理结构设计之后, 数据库设计人员要用 RDBMS 提供的数据库定义语言和其他应用程序将数据库逻辑设计和物理设计结果严格描述出来, 成为 RDBMS 可以接受的源代码, 再经过调试产生目标模式, 然后组织数据入库, 并进行试运行。

6. 数据库运行和维护

数据库应用系统试运行合格后, 就可以投入正式运行了。由于应用环境在不断变化, 数据库运行过程中物理存储也会不断变化, 对数据库设计进行评价、调整、修改等维护工作是一个长期的任务, 也是数据库设计工作的继续和提高。

1.2 思考与练习

1.2.1 选择题

- 下列关于数据库系统的叙述中, 正确的是()。
 - 数据库管理系统由数据库系统、用户、数据库和数据库应用系统组成
 - 数据库管理系统是用户与数据库之间的接口
 - 采用数据库技术完全消除了数据冗余
 - 采用数据库技术降低了数据共享性和独立性
- 下列不属于数据库系统三级模式结构的是()。
 - 外模式
 - 模式
 - 内模式
 - 关系模式

3. 按照数据的组织形式, 逻辑数据模型可分为三种模型, 它们是()。
- A. 独享、共享和实时 B. 网状、环状和链状
C. 层次、网状和关系 D. 概念、逻辑和物理
4. 下列有关数据库的叙述中, 正确的是()。
- A. 在数据库系统中, 数据的物理结构必须与逻辑结构一致
B. 数据库设计是指对数据库系统基础的数据模型的设计
C. 数据库是存储在计算机存储设备中的、结构化的相关数据的集合
D. 数据库系统不需要操作系统支持也可以使用
5. 下列有关数据库的叙述中, 正确的是()。
- A. 数据库是一个关系
B. 数据库是一组文件
C. 数据处理是将信息转化为数据的过程
D. 如果一个关系中的属性或属性组不是本关系的主码, 但它是另一个关系的主码, 则称其为本关系的外码
6. 关系数据库管理系统能实现的专门关系运算包括()。
- A. 增加、删除、更新 B. 选择、投影、连接
C. 关联、更新、排序 D. 索引、统计、汇总
7. 关系数据库中所谓的“关系”是指()。
- A. 表中的两个字段有一定的关系
B. 某两个数据库文件之间有一定的关系
C. 记录中的数据彼此之间有一定的关联关系
D. 数据模型符合满足一定条件的二维表格式
8. 数据模型的三要素不包括()。
- A. 数据查询 B. 数据结构 C. 数据操作 D. 数据约束
9. 现实世界中的事物个体在概念世界中称为()。
- A. 记录 B. 实体 C. 实体集 D. 元组
10. 关系数据库的数据和更新操作必须遵循的完整性规则是()。
- A. 实体完整性和参照完整性
B. 参照完整性和用户自定义完整性
C. 实体完整性和用户自定义完整性
D. 实体完整性、参照完整性和用户定义的完整性
11. 设有如下关系表 R、S 和 T:

R		
A	B	C
1	1	2
2	2	3

S		
A	B	C
3	2	1

T		
A	B	C
1	1	2
2	2	3
3	2	1

- 则下列操作正确的是()。
- A. $T=R \cap S$ B. $T=R \cup S$ C. $T=R-S$ D. $T=R \times S$
12. 应用数据库的主要目的是为了解决()。
- A. 数据的保密问题 B. 数据完整性问题
C. 数据量大的问题 D. 数据共享问题
13. 一家书店的店主想将 Book 表的“书名”设为主键,考虑到有重名的书,但相同书名的作者都不同。若按照店主的需求定义 Book 表的主键,则可以选择()。
- A. 不定义主键
B. 定义自动编号主键
C. 将书名和作者组合定义多字段主键
D. 再增加一个内容无重复的字段定义为单字段主键
14. 下列关于关系数据库文件中各条记录顺序的叙述,正确的是()。
- A. 前后顺序不能任意改变,一定要按照关键字段值的顺序排列
B. 前后顺序不能任意改变,一定要按照输入的顺序排列
C. 前后顺序可以任意改变,不影响数据库中数据的数据关系
D. 前后顺序可以任意改变,但排列顺序不同,统计处理的结果有可能不同
15. 一支球队由一名主教练、一名队医和若干球员组成,则球队和主教练是()联系。
- A. 一对一 B. 一对多 C. 多对一 D. 多对多
16. 在关系数据库中,主码标识元组通过()实现。
- A. 用户自定义完整性 B. 参照完整性
C. 实体完整性 D. 值域完整性
17. 反映主键与外键之间引用规则的是()。
- A. 用户自定义完整性 B. 参照完整性
C. 实体完整性 D. 关系模型
18. 下列是关系模型的性质描述,错误的是()。
- A. 关系中不允许存在两个完全相同的记录
B. 任意的一张二维表就是一个关系
C. 关系中元组的顺序无关紧要
D. 关系中列的次序可以任意交换
19. 在关系模型中,主码可由()。
- A. 至多一个属性组成
B. 一个或多个其值能唯一标识该关系模型中任何元组的属性组成
C. 多个任意属性组成
D. 一个或多个任意属性组成
20. 有一张学生表:学生(学号,姓名,性别,年龄,身份证号),则此学生表的候选码是()。
- A. 学号,身份证号 B. 学号,姓名
C. 学号,性别 D. 姓名,身份证号

21. 若有表示学生选课的一张表: 学生(学号, 姓名, 性别, 年龄, 身份证号), 课程(课号, 课名), 选课(学号, 课号, 成绩), 则选课表的关键字(也称键或码)是()。
- A. 课号, 成绩 B. 学号, 成绩
C. 学号, 课号 D. 学号, 姓名, 成绩
22. 要显示 Stu 表中学生姓名和性别的信息, 应采用的关系运算是()。
- A. 选择 B. 投影 C. 连接 D. 交叉
23. 要显示 Stu 表中所有女学生的信息, 应采用的关系运算是()。
- A. 选择 B. 投影 C. 连接 D. 交叉
24. 将两个关系拼接成一个新的关系, 生成的新关系中包含满足条件的元组, 这种操作称为()。
- A. 选择 B. 投影 C. 连接 D. 笛卡尔积
25. 在数据库应用系统开发过程中, 需求分析阶段的主要任务是确定系统的()。
- A. 系统功能 B. 数据模型
C. 开发费用 D. 开发技术
26. 在数据库设计中, 将 E-R 图转换成关系数据模型的过程属于()阶段。
- A. 需求分析 B. 概念设计
C. 逻辑设计 D. 物理设计
27. 数据库与文件系统的根本区别是()。
- A. 提高了系统效率 B. 数据的结构化与共享
C. 节省了存储空间 D. 方便了用户使用
28. 在开发企业进销存管理系统过程中到企业调研, 属于数据库应用系统设计中()阶段的任务。
- A. 物理设计 B. 概念设计
C. 逻辑设计 D. 需求分析
29. 数据库系统的数据独立性体现在不会因为()。
- A. 数据的变化而影响到应用程序
B. 系统数据存储结构与数据逻辑结构的变化而影响应用程序
C. 存储策略的变化而影响存储结构
D. 某些存储结构的变化而影响其他的存储结构
30. 关系数据规范化的意义是()。
- A. 保证数据的安全性和完整性
B. 提高查询速度
C. 减少数据操作的复杂性
D. 消除关系数据的插入、删除和修改异常以及数据冗余

1.2.2 填空题

1. _____是数据库中存储的基本对象, 是描述事物的符号记录。
2. 依据_____来划分数据库的类型。数据库的性质由其采用的_____所决定。

3. 若有如下两个关系:

患者(患者编号, 患者姓名, 性别, 出生日期, 职业, 既往病史)

医疗(患者编号, 医生编号, 医生姓名, 诊断日期, 诊断结果)

其中, 医疗关系的主码是_____, 外码是_____。

4. 如果一个护士管理多个病房, 一个病房只被一个护士管理, 则病房与护士之间存在_____联系。

5. 数据独立性高是数据库系统的特点之一, 数据独立性包括_____和_____两种。当数据的逻辑结构改变时, 用户的应用程序可以不改变, 即是指数据具有_____独立性。

6. 在数据库设计过程中, _____和_____阶段的设计与选用的数据库管理系统密切相关。

7. 一个工人可以加工多种零件, 每一种零件可以由不同的工人来加工, 工人和零件之间为_____联系。

8. 关系中的属性或属性组合, 其值能够唯一标识一个元组, 该属性或属性组合可选作_____。

9. 在数据库的概念结构设计中, 常用的描述工具是_____。

10. 数据库管理系统是位于_____之间的软件系统。

11. 关系运算是针对关系数据库的数据操纵, 主要用于关系数据库的_____操作。

12. 按照运算符的不同, 关系代数的运算可分为_____和_____两类。

13. 由 2023 年 3 月组建的_____等 17 个部门联合发布《“数据要素×”三年行动计划(2024—2026 年)》。

14. 目前蚂蚁集团、网商银行的全部核心系统都由_____数据库支撑。

15. 2021 年 6 月 10 日, 第十三届全国人民代表大会常务委员会第二十九次会议通过《中华人民共和国_____法》, 此法规范数据处理活动, 保障数据安全, 促进数据开发利用, 保护个人、组织的合法权益, 维护国家主权、安全和发展利益。

16. 我国著名科学家钱学森在 1990 年将虚拟现实技术的元宇宙翻译为_____。

17. 实体与实体之间的联系有_____、_____和_____3 种。

18. “顾客”与“商品”两个实体集之间的联系一般是_____。

19. 一个关系表的行称为_____或_____。

20. 自然连接是_____的等值连接。

1.2.3 简答题

1. 数据管理技术经历了哪三个阶段? 请简述各阶段的特点。
2. 数据库系统具有哪些特点?
3. 数据冗余可能引起哪些问题?
4. 数据库管理系统有哪些主要功能? 请列举几个常见的数据库管理系统。
5. 什么是数据模型? 数据模型应满足哪些要求? 数据模型按不同的应用层次分为哪

三种类型？

6. 层次模型、网状模型和关系模型的数据结构是什么？请简述它们的优缺点。
7. 关系模型有什么特点？请简述关系模型的主要术语。
8. 超市每个时段要安排一个班组上岗值班，每个收银口要配备两名收银员配合工作，共同使用一套收银设备为顾客服务。请分析“顾客”与“收银口”、“班组”与“收银员”、“收银口”与“收银设备”、“收银口”与“收银员”的关系。
9. 图 1-1 所示的 E-R 模型中有哪几个实体？每个实体的候选码是什么？码是什么？实体之间有哪几种联系？此 E-R 图反映的课程编排规则是什么？

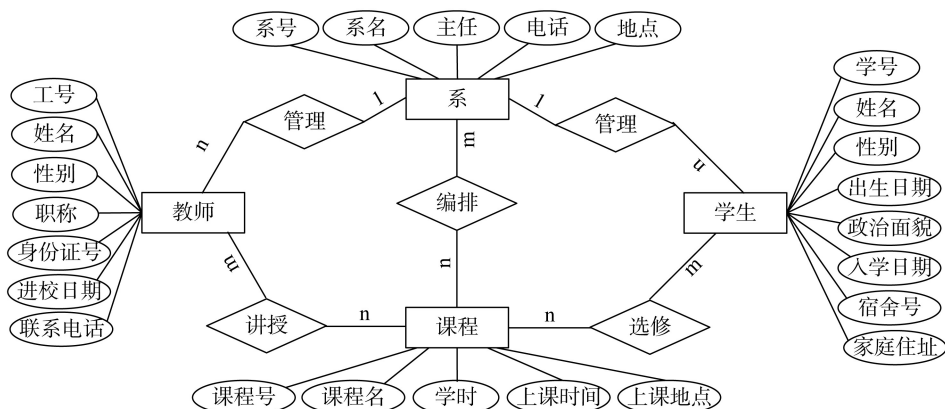


图 1-1 系部排课 E-R 图

10. 请举例说明关系的两个不变性。(实体完整性和参照完整性是所有关系模型必须满足的数据完整性约束，被称作是关系的两个不变性。)
11. 传统的集合运算包括哪些？如何用差运算来实现交运算？
12. 等值连接和自然连接有什么不同？
13. 数据库设计过程依次分为哪 6 个阶段？前 4 个阶段的成果分别是什么？
14. 在数据库设计的逻辑结构设计阶段，E-R 图向关系模型转换应遵循什么原则？
15. 试用关系模式的规范化理论分析表 1-1 中存在的问题，并手工分解成符合范式要求的模式。

表 1-1 机动车驾驶证申请条件汇总表

准驾车型	是否初学	身体条件		年龄条件		增驾条件	可否在暂住地申请
		身高 (cm)	视力	申请年龄	允许年龄	驾驶经历及记分情况	
A1	否	155	5.0	26~50	26~60	B1、B2 五年以上且前三个周期内无满分记录 A2 两年以上且前一个周期内无满分记录 无死亡事故中负主要以上责任的记录	不可
...

1.3 实验案例

实验案例 1

案例名称：创建高校教学系统的实体-联系模型

【实验目的】

掌握用 E-R 图方法表示概念模型。

【实验内容】

本实验完成以下两项任务：

(1) 依据所述情况创建概念模型。某高校有若干个学院，每个学院有若干专业和教研室，每个教研室有若干教员，其中有教授或副教授职称的教员可带若干研究生。每个专业有若干班级，每个班有若干学生，每个学生选修若干课程，每门课可由若干名学生选修。

(2) 请到教务处调研高校教学排课管理业务，设计排课概念模型。参考实体：学生、课程、教师、教室。

请在 Microsoft Office Visio 中画出概念模型的 E-R 图。

实验案例 2

案例名称：创建大学生创新创业训练计划的概念模型

【实验目的】

掌握用 E-R 图或 UML 方法创建概念模型以解决实际问题。

【实验内容】

国家级大学生创新创业训练计划内容包括创新训练项目、创业训练项目和创业实践项目三类。

创新训练项目是本科生个人或团队，在导师指导下，自主完成创新性研究项目设计、研究条件准备和项目实施、研究报告撰写、成果(学术)交流等工作。

创业训练项目是本科生团队，在导师指导下，团队中每个学生在项目实施过程中扮演一个或多个具体的角色，参与编制商业计划书、开展可行性研究、模拟企业运行、参加企业实践、撰写创业报告等工作。

创业实践项目是学生团队，在学校导师和企业导师共同指导下，采用前期创新训练项目(或创新性实验)的成果，提出一项具有市场前景的创新性产品或者服务，以此为基础开展创业实践活动。

请从三类创新创业训练计划中任意选取一种，确定实体、实体的属性及实体间的联系，建立概念模型。

实验案例 3

案例名称：创建图书管理数据库的关系模式

【实验目的】

掌握 E-R 图向关系模型转换的规则，尝试以规范化理论为指导对关系模型进行优化。

【实验内容】

设计一个图书管理数据库，用 E-R 图画出它的概念模型，再将其用关系模式表示。此图书管理数据库中，每本图书的信息包括：书号、书名、作者、出版社和出版日期；每本被借图书的信息包括：读者编号、借出日期和应还日期；每位借阅者的信息包括：读者编号、姓名、性别、单位、已借阅数、最大可借阅数和违规记录。

为了显示更加直观，关系模式中的主键名请用下画线标出，外键名请用斜体表示。

实验案例 4

案例名称：大学生竞赛管理系统的逻辑数据模型设计

【实验目的】

联系实际，通过调查分析，设计逻辑数据模型，为在 Access 2016 中实现数据库应用系统做准备。

【实验内容】

大学生竞赛管理系统主要包括组织学生报名参赛、设置竞赛场地和场次、聘请评审专家、设计赛事议程和评审指标、设置奖项与奖品颁发等活动。请围绕这些活动，依据关系数据库理论设计逻辑数据模型。

大学生竞赛是通过各高校组队报名的形式组织学生参赛，一所院校可报多个团队，一个团队限报一件作品，一个学生可以参加多个团队，一个团队可由不超过 3 个学生组成，一件作品由多个专家评审，一个专家可以评审多件作品。由此可以设计：参赛学生应有学号、姓名、报名号、专业、年级、所属团队等属性；作品有作品编号、作品名称、作品类别、制作日期、作品简介、作品效果图、指导老师、参赛人数、报名号、赛场 ID 等属性；报名表有报名号、院校名、所属地区、组队、缴纳报名费等属性；赛场有赛场 ID、赛场名称、赛场地点、竞赛时间等属性；专家有专家号、专家姓名、职务职称、专业、单位、联系电话等属性。

请进一步细化以上各实体的属性，确定主码和外码。参赛作品与参赛学生、评审专家之间都有联系，如何建立这些联系，以构成一个较完整的竞赛管理逻辑数据模型？请按数据库设计流程完成设计。

实验案例 5

案例名称：高校学生社团管理系统的逻辑数据模型设计

【实验目的】

联系实际，通过调查分析，设计逻辑数据模型，为在 Access 2016 中实现数据库应用

系统做准备。

【实验内容】

大学生校园文化丰富多彩,校方鼓励在校学生创办、参加各类社团。为加强社团管理,校团委成立社团联合会(社联)对学生社团进行管理。高校学生社团管理系统的使用者主要是大学生、社团、社联及校团委。

大学生通过社团管理系统浏览社团简介、社团活动信息,以及一些通知公告或者招募信息,在社团/社联纳新时可以申请加入社团/社联,在达到一定的条件之后可以申请成立新社团。

社团的主要职能是申请活动、举办活动、管理学生的进团和退团。社团中包括负责人、社团成员和社团部门。

社联的主要职能是管理各个学生社团,审批社团的活动申请,组织社联活动,对学生进出社联进行管理,社联中包括负责人、社联成员和社联部门。

校团委要对社联/社团进行工作指导和监管,一些大型的社联/社团活动需要校团委审批。

请结合自身加入社团/社联的经历,设计高校学生社团管理系统的逻辑模型。

实验案例 6

案例名称:基础电信业务逻辑数据模型设计

【实验目的】

通过调查分析,设计基础电信业务逻辑数据模型。

【实验内容】

电信业务是指电信网向公众提供的业务。电信业务根据业务类型分为基础电信业务和增值电信业务。基础电信业务又分为第一类、第二类两种。第一类基础电信业务是指固定通信、移动通信、卫星通信和数据通信;第二类基础电信业务是指集群通信、无线寻呼、卫星通信、数据通信、网络接入、设施服务和网络托管。

固定通信是指通信终端设备与网络设备之间主要通过电缆或光缆等线路固定连接,进而实现用户间相互通信,其主要特征是终端的不可移动性或有限移动性,如普通电话机、IP 电话终端、传真机、无绳电话机、联网计算机等电话网和数据网终端设备。固定电信业务包括:固定网本地电话业务、固定网国内长途电话业务、固定网国际长途电话业务、IP 电话业务和国际通信设施服务业务等。

下面分别给出电信业务的客户资料表和客户出账表的常用字段。客户资料表:客户标识、客户类别、客户姓名、电话号码、证件类型、客户证件号码、付费方式、入网日期等;客户出账表:客户标识、基本月租费、增值服务费、本地通话费、长途通话费、总费用等。

请以第一类或第二类基础电信业务为建模对象,构建电信业务逻辑模型。

实验案例 7

案例名称:旅游管理信息系统模型设计

【实验目的】

通过调查分析，设计旅游管理信息系统数据模型。

【实验内容】

旅游管理信息系统中与业务有关的信息应包含：旅游线路、旅游班次、旅游团、游客、保险、导游、宾馆、交通工具等。“旅游线路”包括线路号、起点和终点等属性；“旅游班次”包括班次号、出发日期、天数和报价等属性；“旅游团”包括团号、团名、人数、联系人等属性；“游客”包括身份证号、姓名、性别、年龄、电话等属性；“导游”包括导游证号、姓名、性别、电话、等级等属性；“宾馆”包括宾馆编号、宾馆名称、星级、房价、电话等属性；“交通工具”包括车次、车型、座位数、司机姓名等属性；保险单包括保单号、保险费、投保日期等属性。

请到所在地的旅游机构进行调研，完善以上内容，画出旅游管理信息系统 E-R 图，并将其转化为关系模式。若两实体之间是多对多联系，请将其转化为两个关系模式。

实验案例 8

案例名称：汽车运输公司运营模型设计

【实验目的】

通过调查分析，设计汽车运输公司运营数据模型。

【实验内容】

汽车运输公司运营数据库中有 3 个实体集：“车队”实体集有车队编号、车队名称、车队负责人等属性；“司机”实体集有司机工号、姓名、性别、年龄、电话等属性；“车辆”实体集有牌照号、车型、出厂日期等属性。规则：车队与司机之间存在“聘用”联系，每个车队可聘用若干司机，但每个司机只能被一个车队聘用，车队聘用司机有聘期；车队与车辆之间存在“拥有”联系，每个车队可拥有若干车辆，但每辆车只能归属一个车队；司机与车辆之间存在“驾驶”联系，司机驾驶车辆有驾驶日期、公里数、违章记录等属性，每个司机可以使用多辆汽车，每辆汽车可被多个司机使用。

请到所在地的汽车运输公司进行调研，完善以上内容，画出注明属性、联系类型和实体的 E-R 图，并将其转化为关系模式，标注主键和外键。若两实体之间是多对多联系，请将其转化为两个关系模式。

实验案例 9

案例名称：商业集团数据库管理系统模型设计

【实验目的】

依据关系数据库理论，设计商业集团数据库管理系统数据模型。

【实验内容】

某商业集团数据库有 5 个实体集：“公司”实体集有公司编号、公司名、法定代表人、注册资金、地址等属性；“职工”实体集有职工编号、姓名、性别等属性；“仓库”实体集

有仓库编号、仓库名、地址等属性；“商店”实体集有商店号、商店名、店长、地址等属性；“商品”实体集有商品号、商品名、单价等属性。

公司与仓库之间存在“隶属”联系，每个公司管辖若干仓库，每个仓库只能被一个公司管辖；仓库与职工之间存在“聘用”联系，每个仓库可聘用多个职工，每个职工只能在一个仓库工作，仓库聘用职工有聘期和工资；仓库与商品之间存在“库存”联系，每个仓库可存储若干种商品，每种商品存储在若干仓库中，每个仓库在存储一种商品时登记存储日期和存储数量；商店与商品之间存在“销售”联系，每个商店可销售若干种商品，每种商品可在若干个商店里销售，每个商店在销售一种商品时登记月份和月销量；仓库、商店、商品之间存在“供应”联系，有月份和月供应量两个属性。

请画出商业集团数据库管理系统 E-R 图，在图上注明属性和联系的类型；将 E-R 图转换为关系模型，并注明主码和外码。

【思考】

大型连锁超市信息管理系统的关系模型如何设计？

实验案例 10

案例名称：关于电影的数据库模式分析

【实验目的】

依据关系数据库理论，对给出的 5 个关系模式进行分析。

【实验内容】

电影 Movies(title,year,length,genre,studioName,producer#)

电影明星 MovieStar(name,address,gender,birthdate)

演出 StarsIn(movieTitle,movieYear,starName)

电影制片 MovieExec(name,address,cert#,netWorth)

电影公司 Studio(name,address,presC#)

试分析：各关系模式属性的含义，各关系的主键是什么？各关系之间存在怎样的联系？

实验案例 11

案例名称：XML 模型应用

【实验目的】

加深理解 XML 模型的特点，学会用以树型结构展示 XML 文档信息。

【实验内容】

(1) 在记事本中录入以下 XML 代码，分别以文件名 XMLcase.txt 和 XMLcase.html 保存。注意文件扩展名。

```
<? xml version="1.0" encoding="UTF-8" standalone="yes"?>
<!-- This is an experimental case -->
<bookstore>
```

```
<book category=" FICTION">
  <title lang="en">Harry Potter</title>
  <author>J K. Rowling</author>
  <year>2005</year>
  <price>28.99</price>
</book>
<book category="DATABASE">
  <title lang="en">Database System Concepts</title>
  <author>Abraham Silberschatz</author>
  <year>2012</year>
  <price>90.00</price>
</book>
<book category="DATAWAREHOUSE">
  <title lang="en">Building the Data Warehouse</title>
  <author>William H.Inmon</author>
  <year>2005</year>
  <price>35.10</price>
</book>
</bookstore>
```

(2) 打开 XMLcase.html 文件，默认打开的应用程序是什么？显示的信息是什么？为什么？

(3) 打开 XMLcase.txt 文件，默认打开的应用程序是什么？显示的信息是什么？

(4) 分析此 XML 文档结构，画一张根元素在顶端的树形图，表示标题为 Harry Potter 的书籍。

树形图可以参照图 1-2。

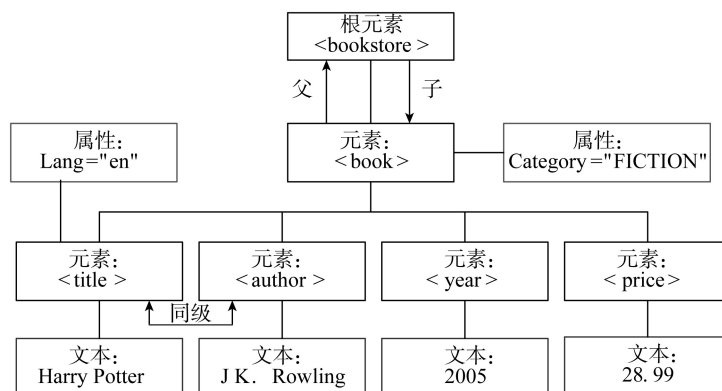


图 1-2 XML 文档树形图

根元素是`<bookstore>`，文档中的所有`<book>`元素都被包含在`<bookstore>`中。`<book>`元素有`<title>`、`<author>`、`<year>`和`<price>`四个子元素。

第 2 章

“人工智能+” 数据技术

2.1 知识要点

2.1.1 人工智能的定义

人工智能(Artificial Intelligence, AI)于 1956 年诞生,像许多新兴学科一样,至今尚无统一定义。

现行国家标准GB/T 11457—2006《信息技术 软件工程术语》对人工智能的定义:计算机科学的一个分支,专门研制执行通常与人的智能有关联的功能(例如,推理、学习和自改进)的数据处理系统;某一设备执行通常与人的智能有关联的功能(例如,推理、学习和自改进)的能力。

百度百科对人工智能的定义:研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。

从宏观层面来看,人工智能可分为弱人工智能、通用人工智能和超级人工智能三个发展阶段。

1. 弱人工智能(Narrow AI)

弱人工智能也称狭义人工智能,是指专门针对特定任务的人工智能系统。例如,语音识别软件、推荐系统、自动驾驶汽车中的某些功能等。这些系统在特定任务上表现出色,但在其他领域则无法应用。

2. 通用人工智能(Artificial General Intelligence, AGI)

通用人工智能也称强人工智能，是指一种能够像人类一样思考、学习和执行多任务的人工智能系统，它具有高效的学习和泛化能力，能根据所处的复杂动态环境自主产生并完成任务，它具备自主感知、认知、决策、学习、执行和社会协作等能力，且符合人类情感、伦理和道德观念。全球生成式 AI 领军者 OpenAI 公司将 AGI 写在了自己的企业使命中。

3. 超级人工智能(Artificial Superintelligence, ASI)

超级人工智能是指在所有领域都远远超越人类智能水平的人工智能系统。这种系统不仅能够执行人类的所有任务，还能在这些任务上展现出前所未有的效率和创造力。ASI 目前还属于科幻和理论探讨的范畴。

2.1.2 人工智能的起源与发展

1. 孕育期(1956 年前)

人类对智能机器和人工智能的梦想和追求可追溯到三千多年前我国西周、东汉和三国时期。20 世纪初，数理逻辑研究取得重大突破，为人工智能的逻辑推理和符号处理奠定了理论基础。1943 年，麦卡洛克和皮茨提出人工神经网络的概念并构建人工神经元的 MP 模型，开创了人工神经网络研究时代。1949 年，唐纳德·赫布出版《行为的组织》，提出 Hebb 学习规则，为机器学习中的神经网络的学习算法奠定基础。1950 年，阿兰·麦席森·图灵发表《计算机与智能》(Computing Machinery and Intelligence)论文，提出“图灵测试”，为判断机器是否具有智能提供了一种方法，被广泛认为是人工智能的开端，图灵被誉为人工智能之父。

2. 诞生与初步发展(1956 年至 20 世纪 60 年代末)

1956 年夏，年轻的数学助教约翰·麦卡锡和他的三位朋友马文·明斯基、纳撒尼尔·罗切斯特和克劳德·香农，邀请艾伦·纽厄尔和赫伯特·西蒙等科学家在美国的达特茅斯(Dartmouth)学院组织了一个夏季学术讨论班，历时两个月。参加会议的是在数学、神经生理学、心理学和计算机科学等领域从事教学和研究工作的学者，在会上第一次正式使用了“人工智能”这一术语，这标志着人工智能学科正式诞生。约翰·麦卡锡通常被认为是人工智能之父。

3. 挫折与调整(20 世纪 60 年代末至 70 年代)

人们发现当时的计算机有限的内存和处理速度不足以解决实际的人工智能问题，也难以建立庞大的数据库帮助程序学习。由于研究进展未达预期，英国政府、美国国防部高级研究计划局等机构逐渐停止了对人工智能研究的资助。

4. 第一次低谷与再次兴起(20 世纪 70 年代至 80 年代末)

20 世纪 70 年代初开始，人工智能研究进入第一次寒冬，科学活动和商业活动衰退，持续近 20 年。20 世纪 80 年代，卡耐基梅隆大学制造出可应用于工业领域的专家系统，企业和大学纷纷参与开发，世界 500 强企业中近一半都研制或使用了专家系统。同时，人工智能数学模型方面取得重大突破，1986 年的多层神经网络和 BP 反向传播算法，推动了神

经网络技术的复兴。

5. 第二次低谷与再次繁荣(20 世纪 80 年代末至 90 年代末)

1987 年至 1993 年,苹果公司、IBM 公司推广的第一代台式机费用远低于专家系统的软硬件开销,且专家系统实用性局限于特定情景,美国国防部高级研究计划局也调整拨款方向,人工智能研究进入第二次寒冬。1997 年,IBM 公司的“深蓝”电脑与国际象棋世界冠军卡斯帕罗夫对战并获得胜利,这是首个电脑系统在标准比赛时限内击败国际象棋世界冠军的事件,展示了人工智能在复杂任务处理上的能力。

6. 稳步发展与应用拓展(2000 年至 2022 年)

21 世纪初,卷积神经网络 CNN 在图像识别任务上表现出色,推动了深度学习的兴起,为自动驾驶、医疗影像等应用奠定了基础。2011 年,IBM 公司的 Watson 在美国智力问答节目上击败两位人类冠军,展现了人工智能在自然语言处理和知识问答方面的能力。人工智能在一些特定领域的应用逐渐成熟。

2017 年 Google 推出 Transformer 模型,为自然语言处理等领域带来重大变革。2018 年 OpenAI 公司开始推出以海量参数和强大生成能力著称的 GPT 系列模型。2022 年生成式人工智能(Generative AI)开始崛起。

7. 广泛应用与深入发展(2023 年至今)

2023 年大模型正式进入开源商用阶段,生成式预训练模型的应用成为人们日常生活中的热门工具,人工智能大规模应用元年到来。2024 年几乎所有重要的模型供应商都发布了多模态模型,人工智能正处于广泛应用与深入发展阶段,同时也在面临伦理、法律和社会影响等方面的挑战。未来,AI 的发展可能会更加注重可解释性、安全性,以及与人类的和谐共生。

2.1.3 人工智能的主要学派

目前人工智能的主要学术流派包括符号主义、连接主义和行为主义三种。这些学派各自拥有独特的理论基础、应用场景和技术手段,共同推动了人工智能技术的发展。

1. 符号主义(Symbolism)

符号主义又称逻辑主义、心理学派或计算机学派。该学派认为人工智能源于数理逻辑,人的认知基元是符号,而且认知过程即符号操作过程。知识是信息的一种形式,是构成智能的基础,人工智能的核心问题是知识表示、知识推理和知识运用。

2. 连接主义(Connectionism)

连接主义又称仿生学派或生理学派,其主要观点是人工智能可以通过模拟人脑的神经系统结构来实现。大脑是由神经元组成的复杂网络,神经元之间通过突触相互连接,智能活动是由大量神经元的集体活动所产生的。

3. 行为主义(Behaviourism)

行为主义又称进化主义或控制论学派,该学派认为智能行为可以在与环境的交互作用中不断进化和学习得到,强调智能是在感知环境和做出行动的循环过程中涌现出来的,而

不是通过内在的符号表示或复杂的神经网络连接。

2.1.4 机器学习

机器学习(Machine Learning)是人工智能的一个重要分支,是一种能够根据输入数据训练模型的系统。它的主要目标是让计算机系统通过对模型进行训练,使计算机能够从新的或以前未见过的数据中得出有用的预测。

机器学习主要分为监督学习、无监督学习、强化学习和联邦学习等类型。

1. 监督学习(Supervised Learning)

监督学习也称有导师学习、有监督学习,它使用已标记的训练数据,即数据集中的每个样本都包含输入特征和对应的输出标签(目标值)。模型的任务是学习输入和输出之间的映射关系,从而能够对新的输入数据进行准确的预测。

监督学习的主要类型是分类和回归。

2. 无监督学习(Unsupervised Learning)

无监督学习也称无导师学习、归纳性学习,它是利用未标记的数据进行学习的机器学习方法。在无监督学习中,数据没有预先定义的标签或目标值,模型需要自己发现数据中的结构、模式和规律。例如,给定一组用户的购物行为数据,无监督学习模型可以发现具有相似购物模式的用户群体,而不需要预先知道这些群体的定义。

无监督学习的主要类型是聚类、数据降维和异常检测。

3. 强化学习(Reinforcement Learning)

强化学习也称再励学习、评价学习或增强学习,它让模型在环境里采取行动,获得结果反馈。模型从反馈里学习,从而能在给定情况下采取最佳行动来最大化奖励或是最小化损失。例如,在一个机器人控制的强化学习场景中,机器人(智能体 agent)在房间(环境 environment)中移动,它的每一个动作(action),如向前走、转弯等,都会得到一个奖励(reward)信号,如到达目标位置得到正奖励,碰撞障碍物得到负奖励,机器人通过不断尝试来学习最优的行动策略,以最快的速度到达目标位置并避免碰撞。又如,在一个游戏环境中,游戏的画面和游戏角色的状态构成环境状态,游戏角色的操作,如跳跃、攻击等,是动作空间,完成任务或获得高分得到正奖励,失败或扣分得到负奖励。

强化学习的主要应用场景是机器人控制、游戏行业和资源管理。

4. 联邦学习(Federated Learning)

联邦学习由 Google AI 团队在 2016 年提出,主要是为了解决移动设备的模型训练问题,在保护用户隐私的同时进行模型协同训练。

联邦学习是指一种多个参与方在保证各自原始私有数据不出数据方定义的可信域的前提下,以保护隐私数据的方式交换中间计算结果,从而协作完成某项机器学习任务的模式。例如,多家医院参与联邦学习,每个医院使用自己的患者数据在本地训练模型,然后将训练后的参数加密发送给一个中心服务器,中心服务器聚合这些参数得到一个更准确的疾病诊断模型,而在这个过程中患者的数据始终没有离开医院。

联邦学习的主要应用场景是隐私敏感和物联网领域。

2.1.5 深度学习

深度学习(Deep Learning)是机器学习的一个子领域，其核心在于使用人工神经网络模仿人脑处理信息的方式，通过层次化的方法提取和表示数据的特征。虽然单层神经网络就可以做出近似预测，但是添加更多的隐藏层可以优化预测的精度和准确性。神经网络由许多基本的计算和存储单元组成，这些单元被称为神经元。神经元通过层层连接来处理数据，并且深度学习模型通常有很多层，因此被称为“深度”学习。深度学习模型能够学习和表示大量复杂的模式，在图像识别、语音识别和自然语言处理等任务中非常有效。

2.1.6 AI 的技术发展方向

孙凝晖院士预测人工智能的技术前沿将朝着以下 4 个方向发展。

1. 多模态大模型

从人类视角出发，人类智能是天然多模态的，人拥有眼、耳、鼻、舌、身、嘴(语言)，从 AI 视角出发，视觉，听觉等也都可以建模为 token 的序列，可采取与大语言模型相同的方法进行学习，并进一步与语言中的语义进行对齐，实现多模态对齐的智能能力。

2. 视频生成大模型

OpenAI 公司于 2024 年 2 月发布文生视频模型 Sora，将视频生成时长从几秒钟大幅提升到一分钟，且在分辨率、画面真实度、时序一致性等方面都有显著提升。Sora 的最大意义是它具备了世界模型(World Models)的基本特征，即人类观察世界并进一步预测世界的能力。世界模型是建立在理解世界的基本物理常识之上，观察并预测下一秒将要发生什么事件。虽然 Sora 要成为世界模型仍然存在很多问题，但可以认为 Sora 具备了画面想象力和分钟级未来预测能力，这是世界模型的基础特征。

3. 具身智能

具身智能指有身体并支持与物理世界进行交互的智能体，如机器人、无人车等，通过多模态大模型处理多种传感数据输入，由大模型生成运动指令对智能体进行驱动，替代传统基于规则或者数学公式的运动驱动方式，实现虚拟和现实的深度融合。因此，具有具身智能的机器人，可以聚集人工智能的三大流派：以神经网络为代表的连接主义，以知识工程为代表的符号主义和控制论相关的行为主义，三大流派可以同时作用在一个智能体，这预期会带来新的技术突破。

4. AI4R

AI4R(AI for Research)已成为科学发现与技术发明的主要范式。由于人工智能大模型具有全量数据，具备上帝视角，借助深度学习的能力，人工智能大模型能够比人向前看更多步数。如果能实现从推断到推理的跃升，人工智能模型就有潜力具备像爱因斯坦那样的想象力和科学猜想能力，这将极大地提升人类科学发现的效率，打破人类的认知边界。

2.1.7 我国 AI 产业发展现状

人工智能正成为发展新质生产力的重要引擎，加速与实体经济的深度融合，全面赋能新型工业化，深刻改变工业生产模式和经济发展形态，将对我国加快建设制造强国、质量强国、网络强国和数字中国发挥重要的支撑作用。

人工智能产业链包括基础层、框架层、模型层和应用层等 4 个部分。其中，基础层主要包括算力、算法和数据；框架层主要指用于模型开发的深度学习框架和工具；模型层主要指大模型等技术和产品；应用层主要指人工智能技术在行业场景的应用。

目前，我国人工智能产业在技术创新、产品创造和行业应用等方面实现快速发展，形成庞大市场规模。伴随以大模型为代表的新技术加速迭代，人工智能产业呈现出创新技术群体突破、行业应用融合发展、国际合作深度协同等新特点，亟需完善人工智能产业标准体系。

2.1.8 我国 AI 标准化体系建设总体要求

1. 坚持创新驱动

优化产业科技创新与标准化联动机制，加快人工智能领域关键共性技术研究，推动先进适用的科技创新成果高效转化成标准。

2. 坚持应用牵引

坚持企业主体、市场导向，面向行业应用需求，强化创新成果迭代和应用场景构建，协同推进人工智能与重点行业融合应用。

3. 坚持产业协同

加强人工智能全产业链标准化工作协同，加强跨行业、跨领域标准化技术组织的协作，打造大中小企业融通发展的标准化模式。

4. 坚持开放合作

深化国际标准化交流与合作，鼓励我国企事业单位积极参与国际标准化活动，携手全球产业链上下游企业共同制定国际标准。

2.1.9 我国人工智能标准体系结构

人工智能标准体系结构包括基础共性、基础支撑、关键技术、智能产品与服务、赋能新型工业化、行业应用和安全/治理等 7 个部分。

2.1.10 我国人工智能治理

中国信息通信研究院发布的《人工智能治理蓝皮书(2024 年)》显示，我国坚持人工智能发展与安全并重，强调国家主导，涵盖顶层设计、法律制度、部门规章和技术标准四大层面，形成了由政府引导、多部门协同、公司部门合作参与的全方位治理格局。我国人工智能治理已形成监管备案、伦理审查和安全框架三个维度相互独立又紧密关联，各有侧重又相辅相成的制度保障体系。

我国已有的立法为人工智能治理奠定了扎实的制度基础。《中华人民共和国网络安全法》《中华人民共和国个人信息保护法》《中华人民共和国数据安全法》三部立法从基础设施、数据要素、自动化决策等方面对人工智能进行了要素治理,《中华人民共和国民法典》《中华人民共和国电子商务法》《中华人民共和国反不正当竞争法》等立法针对性回应了人工智能带来的肖像权侵犯、恶意竞价排序等问题。

2.1.11 国家数据基础设施内涵

数据已成为与土地、劳动力、资本、技术等传统要素并列的新型生产要素。党的二十届三中全会提出“建设和运营国家数据基础设施,促进数据共享”。国家发展和改革委员会、国家数据局、工业和信息化部 2024 年 12 月 31 日印发《国家数据基础设施建设指引》。

国家数据基础设施是从数据要素价值释放的角度出发,面向社会提供数据采集、汇聚、传输、加工、流通、利用、运营、安全服务的一类新型基础设施,是集成硬件、软件、模型算法、标准规范、机制设计等在内的有机整体。国家数据基础设施在国家统筹下,由区域、行业、企业等各类数据基础设施共同构成。网络设施、算力设施与国家数据基础设施紧密相关,并通过迭代升级,不断支撑数据的流通和利用。

2.1.12 国家数据基础设施发展愿景目标

国家数据基础设施是数据基础制度和先进技术落地的重要载体。

在数据流通利用方面,建成支持全国一体化数据市场、保障数据安全自由流动的流通利用设施,形成协同联动、规模流通、高效利用、规范可信的数据流通利用公共服务体系。

在算力底座方面,构建多元异构、高效调度、智能按需、绿色安全的高质量算力供给体系。

在网络支撑方面,构建泛在灵活接入、高速可靠传输、动态弹性调度的数据高速传输网络。

在安全方面,构建整体、动态、内生的安全防护体系。

在应用方面,支持传统行业转型升级,赋能人工智能等新兴产业发展。

总体实现“汇通海量数据、惠及千行百业、慧见数字未来”的美好愿景。

2.1.13 国家数据基础设施总体技术架构

国家数据基础设施具有数据采集、汇聚、传输、加工、流通、利用、运营、安全八大能力。

2.1.14 我国数据库产业发展现状

在新一轮人工智能浪潮驱动下,全球数据库产业变革不断,多强竞争格局逐步形成。得益于国家战略引领,我国数据库产业进入蓬勃发展期和关键应用期。

根据中国通信标准化协会发布的《数据库发展研究报告(2024 年)》,2024 年全球数据库市场规模已突破千亿美元大关,约为 1010 亿美元,企业数量和产品种类也在不断增加,中国数据库市场规模达到 74.1 亿美元(约合人民币 522.4 亿元),占全球的 7.34%。预计到

2028年，中国数据库市场总规模将达到930.29亿元，市场复合年均增长率为12.23%。

2.1.15 我国数据库支撑体系

- (1) 标准方面，我国数据库标准体系日益完善助力产业高质量发展。
- (2) 创新方面，非关系型数据库为重点，我国创新能力日益增强。

2.1.16 我国数据库关键技术发展趋势

随着智能化时代的来临，业务应用场景不断丰富，数据库作为数据基础设施的重要组成部分，呈现出技术融合创新发展、新兴技术逐步应用落地和人工智能与数据库双向赋能的特征。

2.2 思考与练习

2.2.1 选择题

1. 人工智能(Artificial Intelligence, AI)诞生于()年。
A. 1943 B. 1949 C. 1956 D. 2018
2. 从宏观层面来看，人工智能可分为三个发展阶段，不包括()。
A. 弱人工智能 B. 强人工智能
C. 超级人工智能 D. 具身人工智能
3. 2017年 Google 推出()模型，为自然语言处理等领域带来重大变革。
A. Transformer B. LSTM C. RNN D. CNN
4. 机器学习类型不包括()。
A. 监督学习 B. 大模型学习 C. 强化学习 D. 联邦学习
5. 无监督学习的类型不包括()。
A. 聚类 B. 回归 C. 数据降维 D. 异常检测
6. 世界上最早用于人工智能研究的编程语言是()。
A. Fortran B. Golang C. LISP D. C++
7. 我国 AI 治理已形成三个维度相互独立又紧密关联，各有侧重又相辅相成的制度保障体系，这三个维度不包括()。
A. 大语言模型 B. 伦理审查 C. 安全框架 D. 监管备案
8. ()是指一种高度可扩展的数据存储架构，它专门用于存储大量原始数据和衍生数据，这些数据可以来自各种数据源并以不同的格式存在，包括结构化、半结构化和非结构化数据。
A. 索引 B. 湖仓一体 C. 散列 D. 数据湖