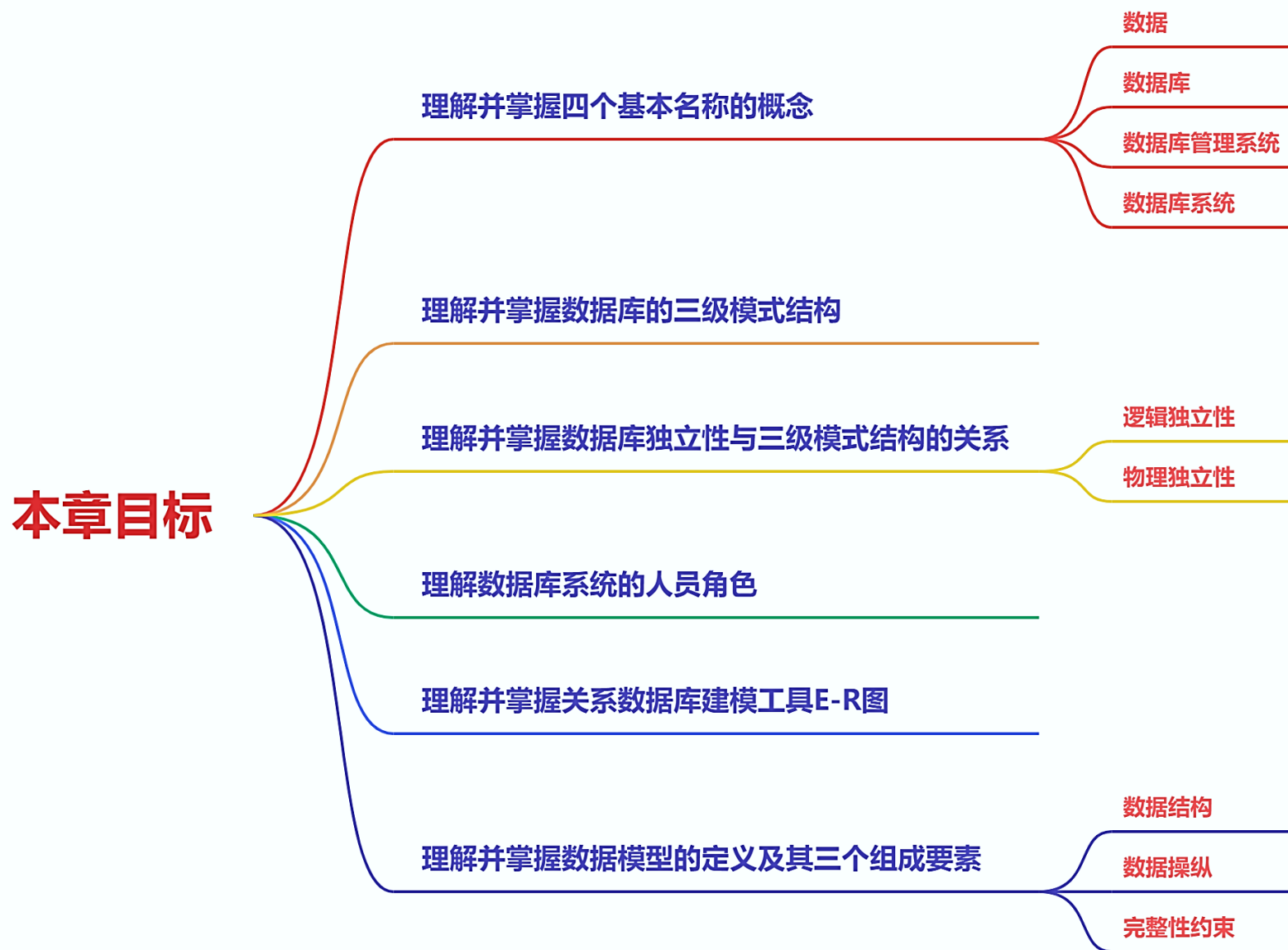


第1章 绪论



- 数据库系统概述
- 数据模型
- 数据库系统三级模式结构
- 数据库系统组成
- 数据库系统体系结构
- 本章小结



1.四个基本概念

- 数据(Data)
- 数据库(Database)
- 数据库管理系统(DBMS)
- 数据库系统(DBS)

2.数据管理技术的产生与发展

- 人工管理阶段、文件系统阶段和数据库阶段

3.数据库系统的特点



- 数据的定义
 - 数据是描述事物的符号记录
- 数据是数据库中存储的基本对象
 - 注意：不仅是存储(静态的观点)，还包括获取、处理、使用、计算等(动态的观点)
- 数据类型
 - 文本、图形、图像、音频、视频、互联网上的博客、微信中的聊天记录、学生的档案记录、个人的网购记录、医院病历等
- 数据的含义称为数据的语义，数据与其语义是不可分的
 - 例子：数据93含义可以是数据库期末总成绩93分|长度93米|摄氏93度|...
 - 学生档案中的学生记录（20230001, 张栋, 男, 2005-3-21, 软件工程）是（学号，姓名，性别，出生日期，主修专业）的一个实例



■ 数据库定义

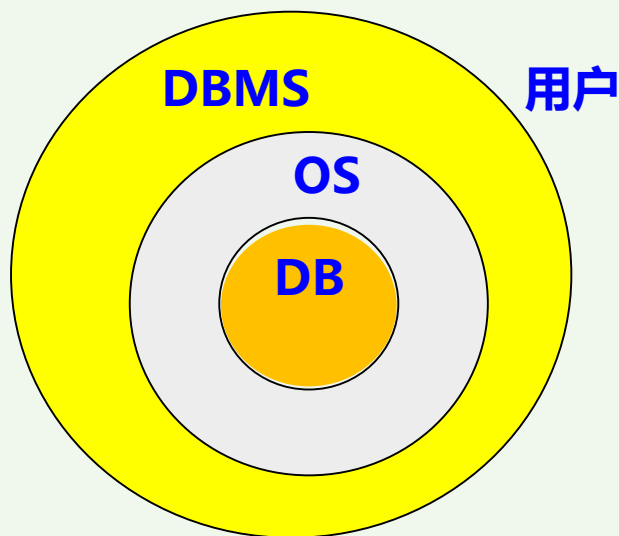
- 是指长期储存在计算机内的、有组织的、可共享的大量数据的集合

■ 数据库的特征

- 数据按一定的数据模型(Data model)组织、描述和存储
- 较小的冗余度
- 较高的数据独立性
- 可扩展性



- Database Management System(简称为DBMS)
 - 位于用户与操作系统之间的一层数据管理软件
 - 是基础软件，是一个大型复杂的软件系统
 - 总是基于某种数据模型，如层次型、网状型、关系型和面向对象型等



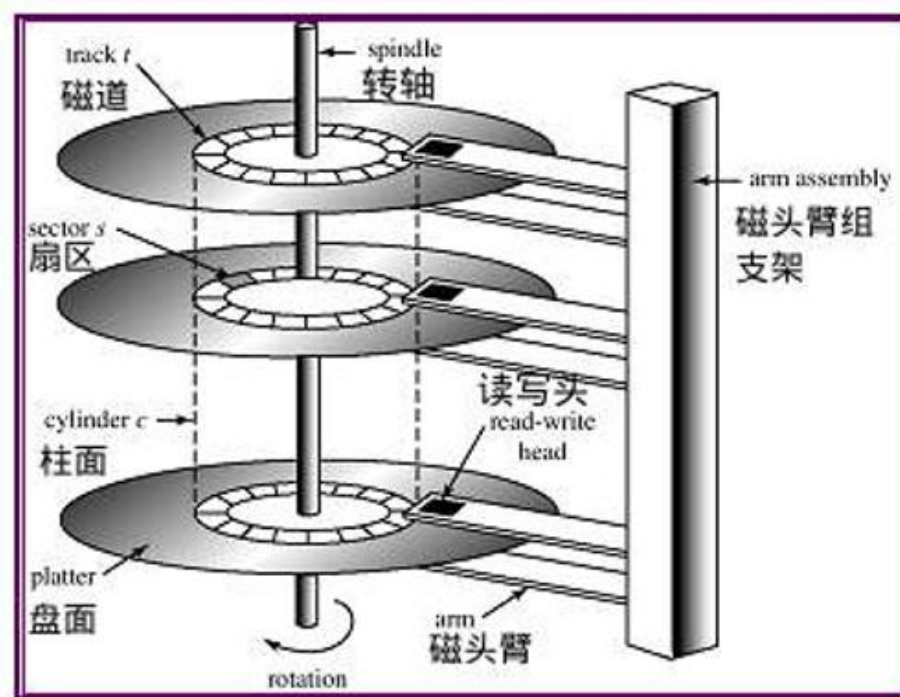
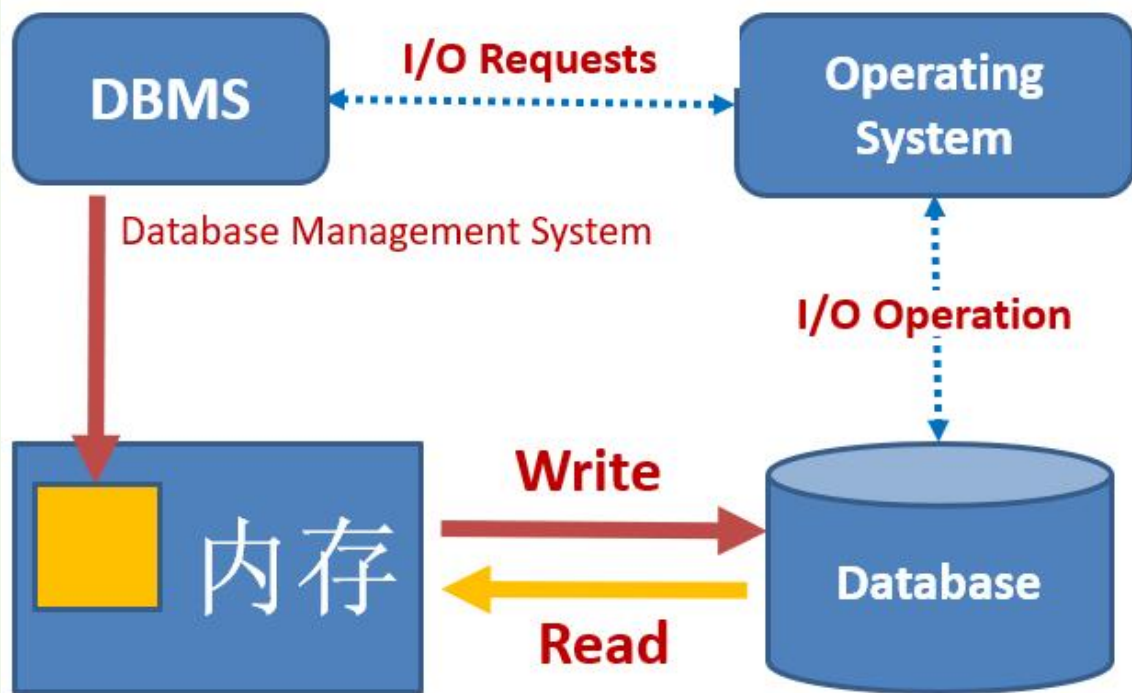


Fig. Conventional Architecture of DBMS



主要功能	主要内容
1.数据定义功能	<ul style="list-style-type: none"> 提供数据定义语言DDL(Data definition language) 定义数据库中数据对象的组成与结构
2.数据组织、存储和管理功能	<ul style="list-style-type: none"> 分类组织、存储和管理各种数据(包括数据字典、用户数据、数据存取路径等) 确定以何种文件结构和存取方式在存储器上组织数据, 如何实现数据之间的联系 目标: 提高存储空间利用率; 存取方便; 提供多种存取方法(如索引技术) 提高存取效率
3.数据操纵功能	<ul style="list-style-type: none"> 提供数据操纵语言DML(Data manipulation language) 实现对数据库的基本操作(查询、插入、删除和修改)
4.数据库事务管理和运行管理功能	<ul style="list-style-type: none"> 数据库在建立、运行和维护时由数据库管理系统统一管理和控制 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
5.数据库建立和维护功能	<ul style="list-style-type: none"> 数据库初始数据的输入和转换 数据库转储和恢复功能 数据库的重组织、性能监视和数据分析等 这些功能通常由一些实用程序或管理工具完成 AI4DB DB4AI
6.其它功能	<ul style="list-style-type: none"> 数据库管理系统与网络中其他软件系统的通信 数据库管理系统系统之间或与文件系统的数据转换 异构数据库之间的互访和互操作



424 systems in ranking, February 2025

Rank			DBMS	Database Model	Score		
Feb 2025	Jan 2025	Feb 2024			Feb 2025	Jan 2025	Feb 2024
1.	1.	1.	Oracle +	Relational, Multi-model ⓘ	1254.82	-3.93	+13.38
2.	2.	2.	MySQL +	Relational, Multi-model ⓘ	999.99	+1.84	-106.67
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model ⓘ	786.87	-11.69	-66.70
4.	4.	4.	PostgreSQL +	Relational, Multi-model ⓘ	659.62	-3.79	+30.21
5.	5.	5.	MongoDB +	Document, Multi-model ⓘ	396.63	-5.87	-23.73
6.	↑ 7.	6.	Redis +	Key-value, Multi-model ⓘ	157.91	+4.55	-2.80
7.	↓ 6.	↑ 9.	Snowflake +	Relational	155.58	+1.68	+28.13
8.	8.	↓ 7.	Elasticsearch	Multi-model ⓘ	134.63	-0.29	-1.11
9.	9.	↓ 8.	IBM Db2	Relational, Multi-model ⓘ	125.44	+2.46	-6.79
10.	10.	10.	SQLite	Relational	113.82	+7.13	-3.47

来源: <https://db-engines.com/en/ranking>

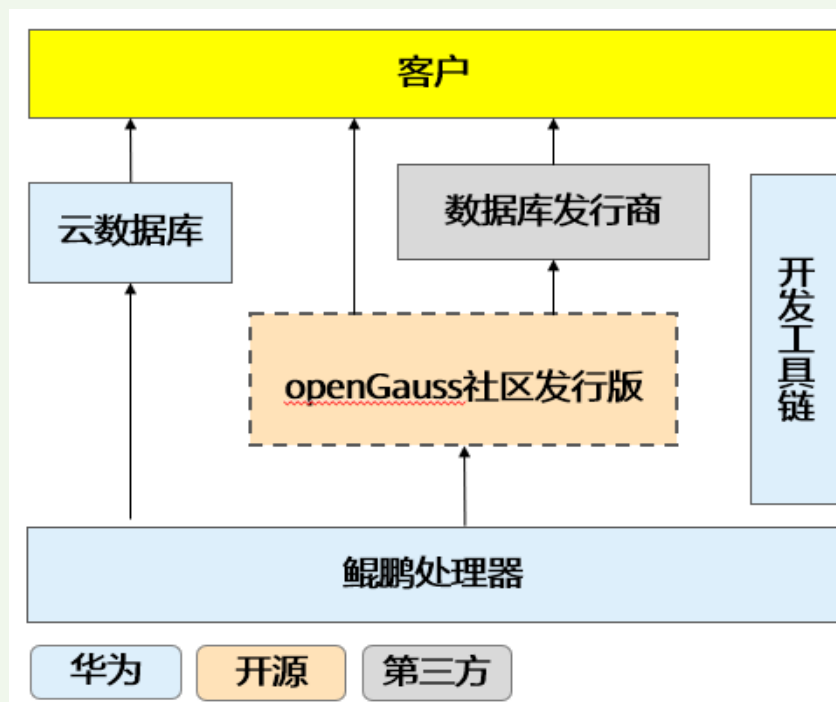
最新解读报告：2025年1月中国数据库排行榜：GaussDB进军三甲雄风起，“OPG”启航迎新年

2025年2月共227个数据库参与，点击查看排名规则更新

排行	上月	半年前	名称	模型 ∨	数据处理 ∨	部署方式 ∨	商业模式 ∨	专利	论文	案例	资质	书籍	得分
	1	↑ 2	OceanBase	关系型	HP	 	 	151	40	55	16	2	692.51
	2	↓ 1	PolarDB	关系型	TP HP	  	 	592	78	85	14	2	686.61
	↑ 4	↑ ↑ ↑ 8	GaussDB	关系型	TP HP	  		950	64	36	17	5	682.74
4	↓ 3	↑ ↑ ↑ 9	GoldenDB	关系型	HP	 		806	54	98	12	2	663.27
5	5	↓ 4	TiDB	关系型	HP	 	 	40	28	19	10	1	587.06
6	6	↓ 5	金仓数据库	关系型	TP AP	  		576	1	64	12	4	570.72
7	7	7	GBASE	关系型	AP TP	  		277	0	84	13	0	551.27
8	8	↓ ↓ 6	达梦数据库	关系型	TP	  		518	0	21	8	12	517.56
9	9	↓ ↓ ↓ 3	openGauss	关系型	TP			570	20	34	7	6	422.66
10	10	10	TDSQL	关系型	HP	 	 	122	23	40	15	0	358.84

来源：<https://www.modb.pro/dbRank>

- 华为openGauss是一款**高性能、高安全、高可靠**的企业级开源关系型数据库
- openGauss集中式版本内核全开源
 - **PG(PostgreSQL)**开源数据库>内部自用孵化阶段>联创产品化阶段>共建生态
- 华为openGauss社区网站：
 - www.opengauss.org



- 支持的硬件平台
 - 支持运行在**ARM**服务器和通用的**x86**服务器上
 - 支持**ARM**服务器和基于**x86_64**的通用**PC**服务器
 - 支持本地存储(**SATA, SAS, SSD**)
 - 支持千兆、万兆**Ethernet**网络
- 支持的操作系统
 - **ARM**
 - **openEuler 20.03LTS** (推荐采用此操作系统)
 - **openEuler 22.03LTS**
 - 麒麟**V10**
 - **Asianux 7.5**
 - **x86**
 - **openEuler 20.03LTS**
 - **openEuler 22.03LTS**
 - **CentOS 7.6**
 - **Asianux 7.6**



技术指标	最大值
数据库容量	受限于操作系统与硬件
单表大小	32TB
单行数据大小	1GB, astore包含CLOB/BLOB类型单行上限为32TB。
每条记录单个字段的大小	1GB, astore包含CLOB/BLOB类型单字段上限为32TB。
单表记录数	最大为 $2^{32} * (8k - \text{页面头}) / \text{行宽}$ 。代码层面的限制是单表最多 2^{32} 个页面，每个页面大小为8k。假设当前数据行宽是1k（包括tuple头），则单表记录数约为 $7 * 2^{32}$ 行（当前页面大小是8k，除了页面头，每个页面包含7行数据）。
单表最大列数	1600（随字段类型不同会有变化，建表时不校验字段类型，存入数据时校验，例如bigint类型的字段，每个字段存入8字节数据，1600个字段，需要存入12800个字节，超过一个页面8K，插入时会报错）
单表中的索引个数	无限制
复合索引包含列数	32
数据库名长度	63
对象名长度（除数据库名以外的其他对象名）	63
单表约束个数	无限制
并发连接数	10000
分区表的分区个数	$2^{20} - 1$ 个
分区表的单个分区大小	32TB
分区表的单个分区记录数	最大为 $2^{32} * (8k - \text{页面头}) / \text{行宽}$ 。代码层面的限制是单表最多 2^{32} 个页面，每个页面大小为8k。假设当前数据行宽是1k（包括tuple头），则单表记录数约为 $7 * 2^{32}$ 行（当前页面大小是8k，除了页面头，每个页面包含7行数据）。
LOB最大容量	(1G - 8203)B
SQL文本最大长度	约为1GB，不同报文接口和处理流程会使用额外空间而略微减少最大可行SQL长度。



■ 标准SQL支持

- 支持标准的SQL92/SQL99/SQL2003/SQL2011规范，支持GBK和UTF-8字符集，支持SQL标准函数与分析函数，支持存储过程

■ 数据库存储管理功能

- 支持表空间，可以把不同表规划到不同的存储位置

■ 提供主备双机

- 事务支持ACID特性、单节点故障恢复、双机数据同步，双机故障切换等

■ 应用程序接口

- 支持标准JDBC 4.0的特性、ODBC 3.5特性

■ 管理工具

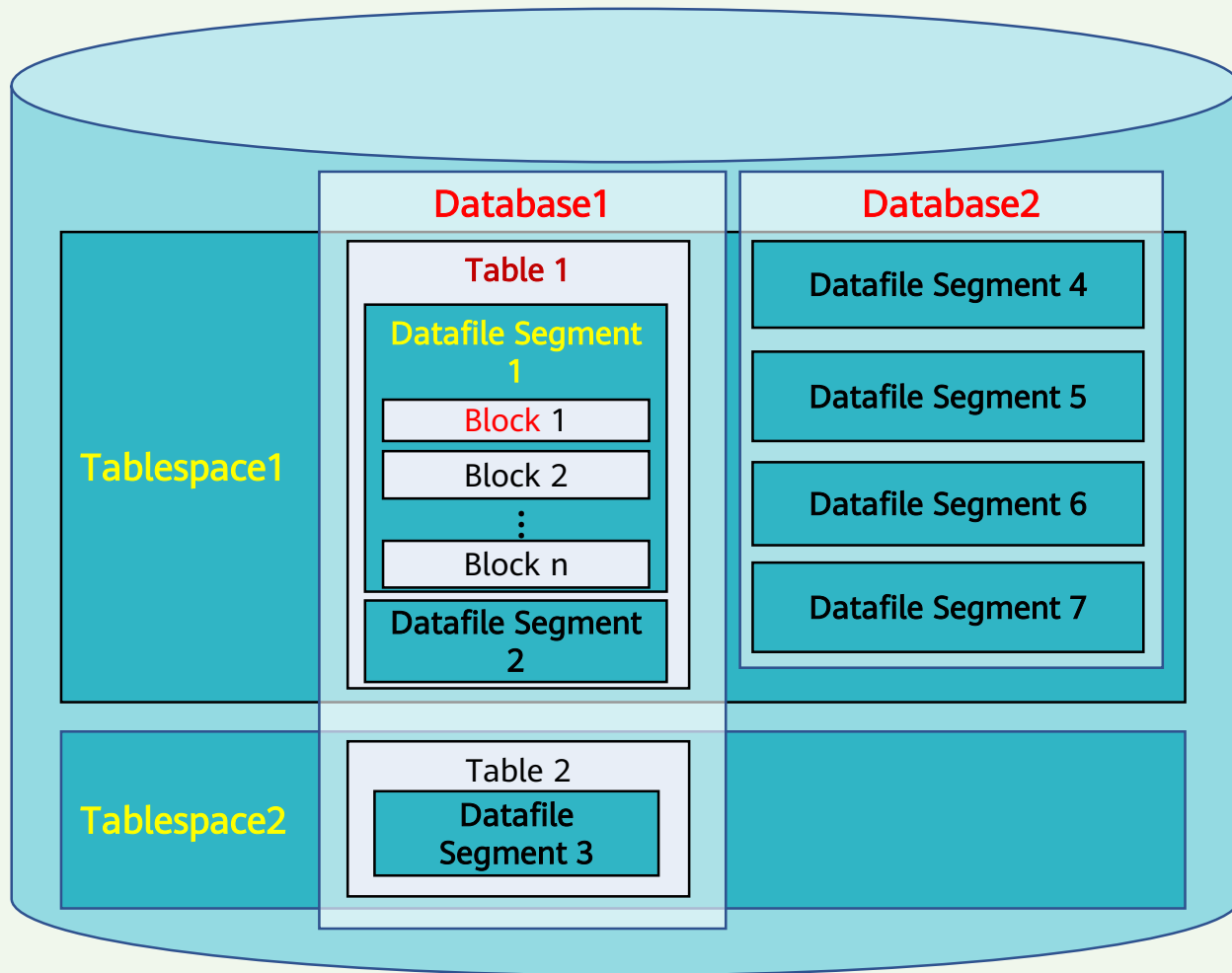
- 提供安装部署工具、实例启停工具、备份恢复工具

■ 安全管理

- 支持SSL安全网络连接、用户权限管理、密码管理、安全审计等功能，保证数据库在管理层、应用层、系统层和网络层的安全性





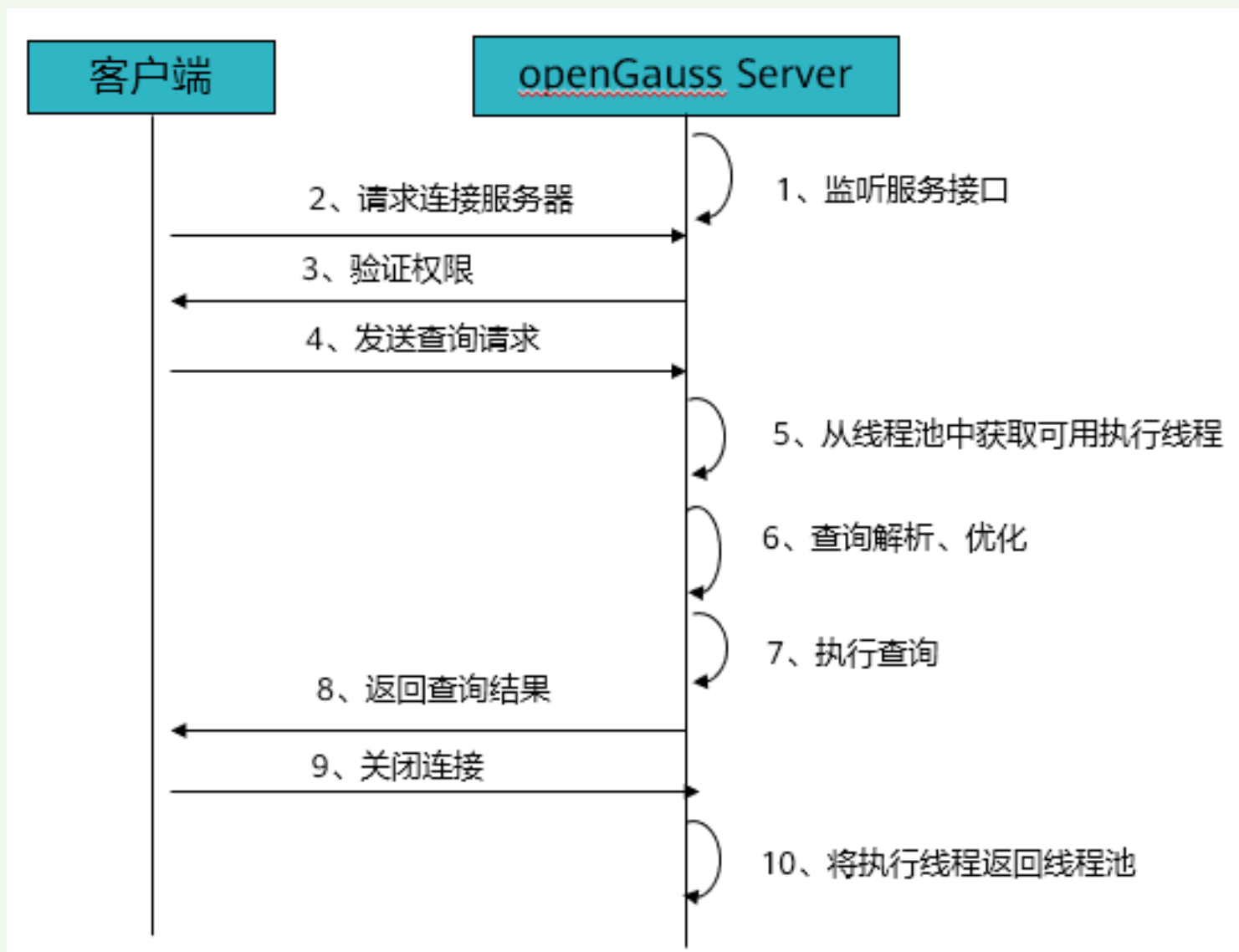


- **Tablespace**: 表空间，是一个目录，可以存在多个，里面存储的是它所包含的数据库的各种物理文件。每个表空间可以对应多个数据库。默认表空间
- **Database**: 数据库，用于管理各类数据对象，各数据库间相互隔离。数据库管理的对象可分布在多个表空间上。
- **Datafile Segment**: 数据文件，通常每张表只对应一个数据文件。如果某张表的数据大于1GB，则会分为多个数据文件存储。
- **Table**: 表，每张表只能属于一个数据库，也只能对应到一个表空间。每张表对应的数据文件必须在同一个表空间中。
- **Block**: 数据块，是数据库管理的基本单位，默认大小为8KB。
- **Segment**: 段，连续的数据块



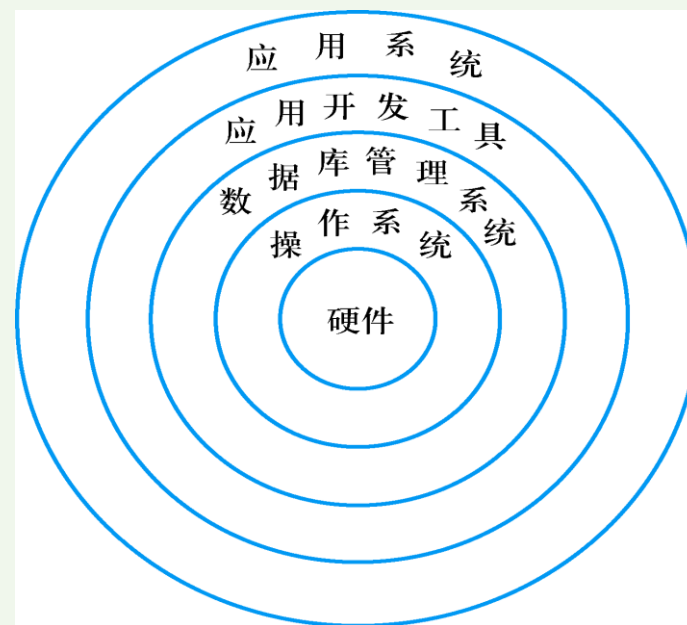
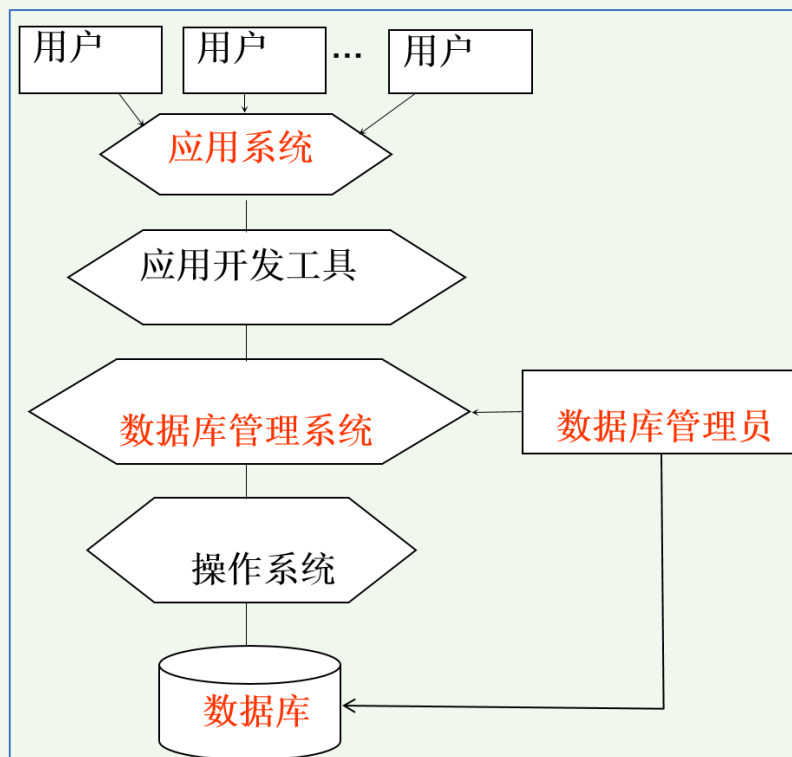
- 数据库的文件默认保存在initdb时创建的数据目录中
- 在数据目录中有很多类型、功能不同的目录和文件
 - 除了数据文件之外，还有
 - 参数文件
 - 控制文件
 - 数据库运行日志
 - 预写日志等





Database System(简称为DBS)

- DBS是指引入数据库后的计算机系统，一般指由数据库、数据库管理系统(及其应用开发工具)、应用系统和数据库管理员(Database administrator, DBA)组成的存储、管理、处理和维持数据的系统



引入数据库后计算机系统的层次机构



1. 四个基本概念

- 数据(Data)
- 数据库(Database)
- 数据库管理系统(DBMS)
- 数据库系统(DBS)

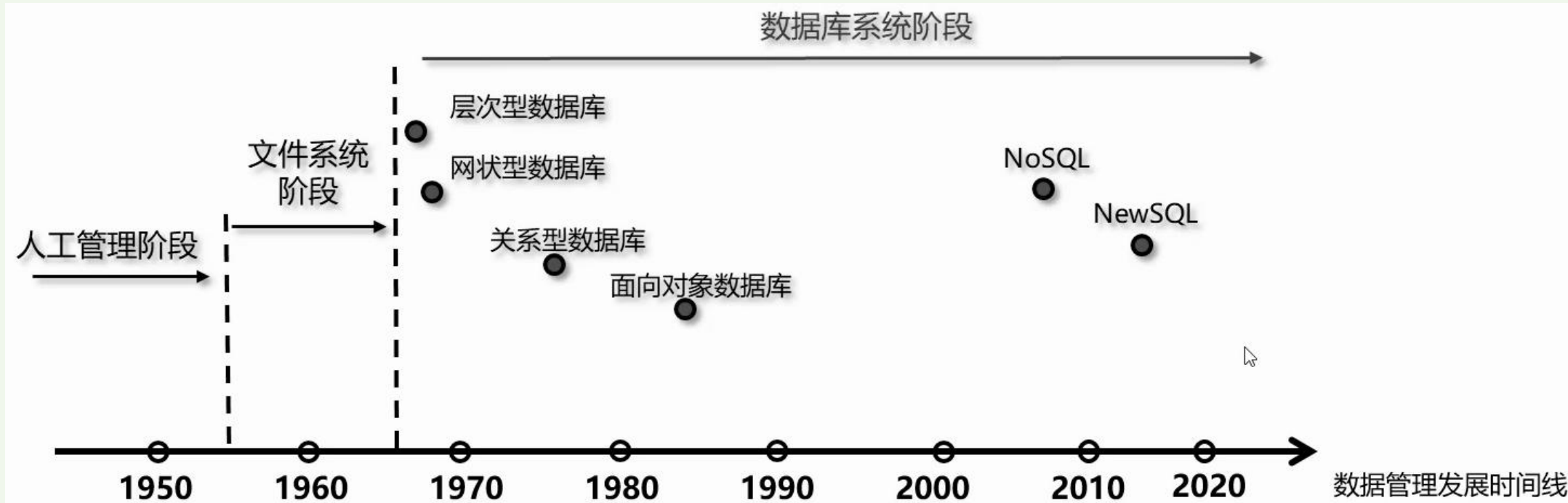
2. 数据管理技术的产生与发展

- 人工管理阶段、文件系统阶段和数据库阶段

3. 数据库系统的特点



- **数据管理**是指对数据进行**分类、组织、编码、存储、检索和维护**等操作，是数据处理的中心问题
- **数据处理**是指对各种数据进行**收集、存储、加工和传播**的一系列活动的总和



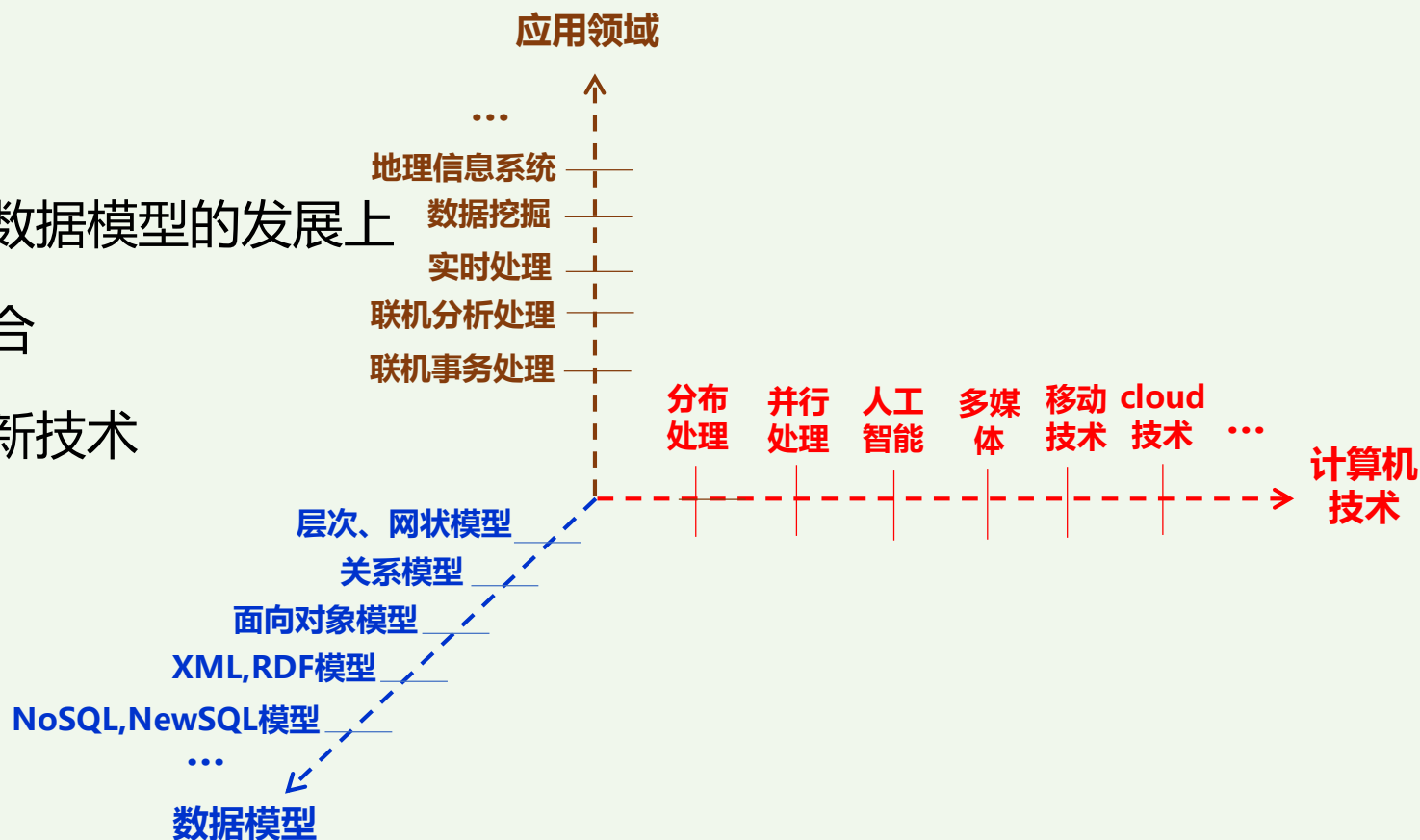
数据管理技术的三个阶段



- 数据库系统已成为计算机信息系统和智能应用系统的核心技术之一和重要基础

- 数据库系统的发展特点

- 数据库的发展集中表现在数据模型的发展上
- 与其它计算机技术交叉结合
- 面向应用领域发展数据库新技术



对比项	人工管理阶段	文件系统阶段	数据库系统阶段
应用领域	科学计算	科学计算、数据管理	大规模数据管理
主要硬件	无直接存取存储设备	磁盘、磁鼓	大容量磁盘、磁盘阵列
主要软件	没有操作系统，没有管理数据的专门软件	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
数据的管理者	人(程序员)	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	现实世界(部门、企业、跨国组织等)
数据的共享程度	不共享，冗余度极大	共享性弱，冗余度高	共享性强，冗余度低且易扩充
数据的独立性	不独立，完全依赖应用程序	独立性弱	具有较强的物理独立性和一定的逻辑独立性
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、并发性控制和数据库恢复功能



1.四个基本概念

- 数据(Data)
- 数据库(Database)
- 数据库管理系统(DBMS)
- 数据库系统(DBS)

2.数据管理技术的产生与发展

- 人工管理阶段、文件系统阶段和数据库阶段

3.数据库系统的特点



1. 整体数据的结构化
2. 数据的共享性强，冗余度低且易于扩充
3. 数据的独立性强
4. 数据由数据库管理系统统一管理和控制

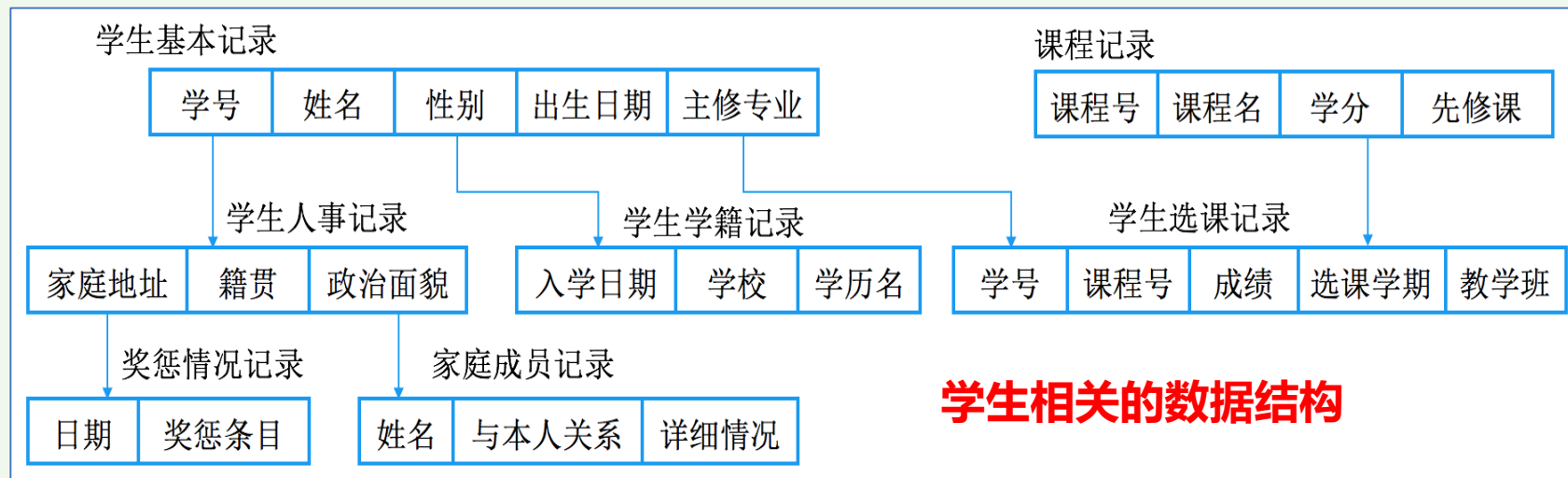


- 整体数据的结构化是数据库的主要特征之一，也是数据库系统阶段与文件系统阶段的本质区别
- 整体数据的结构化含义
 - 指数据库中的数据不再仅仅针对某一个应用，而是面向整个组织或企业的多种应用需求
 - 不仅数据本身结构化，整体是结构化的，数据之间具有联系
 - 即，不仅要考虑某个应用的数据结构，还要考虑整个组织的数据结构
 - 数据结构用数据模型描述，无需程序定义和解释
 - 数据可以变长
 - 数据的最小存取单位是数据项



- 场景：高校本科教务管理信息系统涉及的部门及业务

部门	相关业务
教务处	课程管理、学生选课管理、成绩管理
学生处	学生学籍管理
研究生院	研究生管理
人事处	教师人事管理
科研处	科研管理



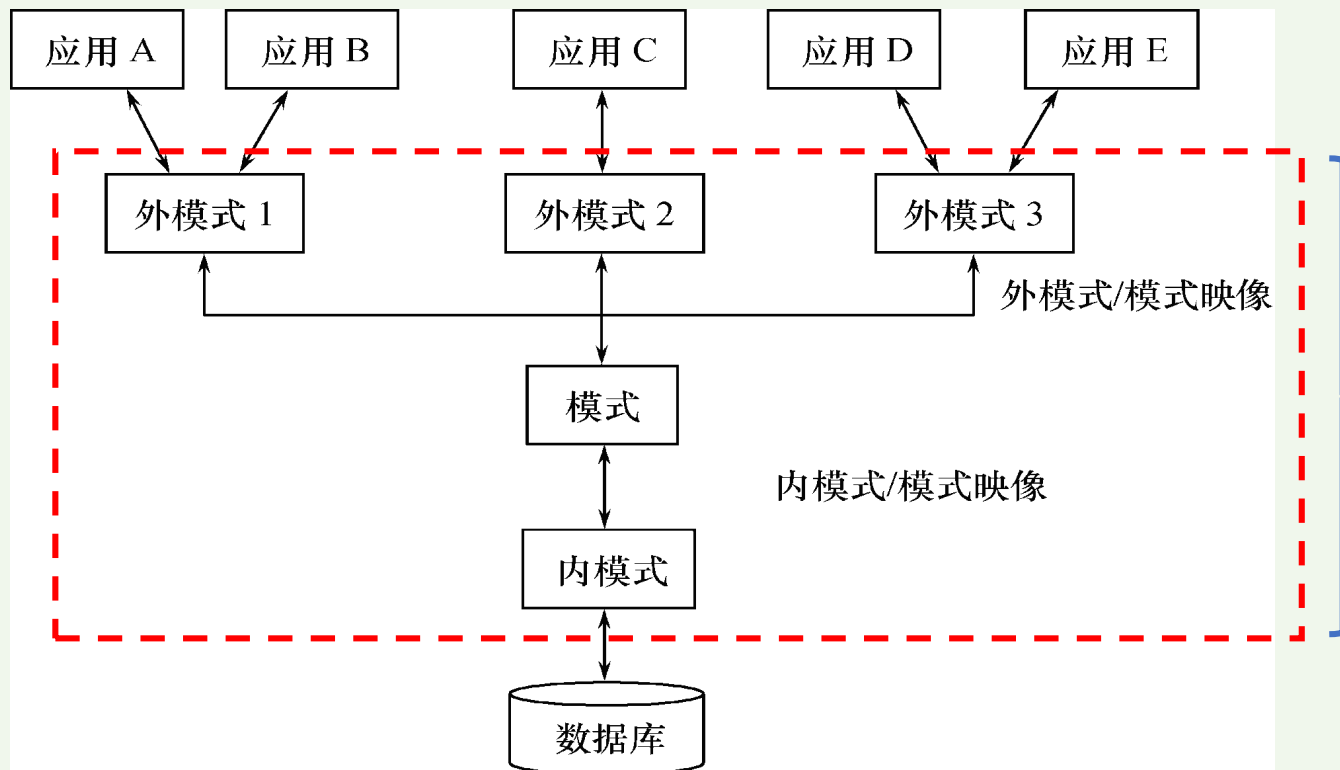
- 数据面向整个系统，可以被多个用户、多个应用、使用不同的接口、不同的编程语言共享使用
- 数据共享的好处
 - 减少数据冗余，节约存储空间
 - 避免数据之间的不相容性与不一致性
 - 使系统易于扩充

指同一数据不同副本的值不一样



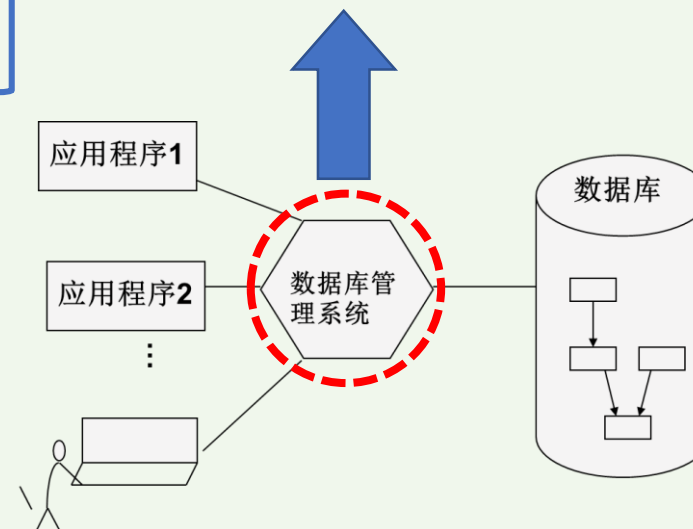
- **数据独立性**是指数据与应用程序相互独立，即数据结构发生变化后，应用程序不必做相应的修改
 - 数据独立性包括**数据的物理独立性**和**数据的逻辑独立性**
 - **数据的物理独立性**
 - 指用户的应用程序与数据库中数据的物理存储是相互独立的
 - 当数据的物理存储改变了，应用程序不用改变
 - **数据的逻辑独立性**
 - 指用户的应用程序与数据库的逻辑结构是相互独立的
 - 数据的逻辑结构改变了，应用程序不用改变
- 数据独立性由数据库管理系统的**两级映像功能**来保证





DBMS的内部架构

三级模式+二级映像



■ DBMS提供的数据控制功能

- 数据的安全性(Security)保护
 - 保护数据，以防止不合法的使用造成数据的泄密和破坏
- 数据的完整性(Integrity)检查
 - 数据的完整性是指数据的正确性、有效性和相容性
 - 将数据控制在有效的范围内，并保证数据之间满足一定的关系
- 并发(Concurrency)控制
 - 对多用户的并发操作加以控制和协调，防止相互干扰而得到错误的结果
- 数据库恢复(Recovery)
 - 将数据库从错误状态恢复到某一已知的正确状态(完整状态或一致状态)



- 数据库是长期存储在计算机内有组织、大量共享数据的集合。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，来保证数据的完整性、安全性，并在多用户同时使用数据库时进行并发控制，在发生故障后对数据库进行恢复
- 数据库系统的出现使信息系统从以加工数据的程序为中心转向以共享的数据库为中心，即以软件为中心向以数据为中心的计算平台的迁移
 - 既便于数据的集中管理，又简化了应用系统的研制和发展，提高了数据的利用率和决策的可靠性



- 模型是对现实世界中某个对象特征的模拟和抽象
 - 例：航模飞机是对生活中飞机的一种模拟和抽象
- 数据模型是对现实世界数据特征的抽象，是现实世界的模拟
 - 具体表现为数据及数据间联系如何表示、数据如何存储、组织和访问
- 数据模型应满足三方面的要求：
 - 较真实地模拟现实世界
 - 易为人理解
 - 便于计算机实现
- 数据模型是数据库系统的核心和基础
 - 现有数据库系统都是基于某种数据模型，如关系、图、向量、文档DB...



- 把现实世界中的具体事物抽象、组织为某一DBMS支持的数据模型的过程称为 **数据建模(Data modeling)**

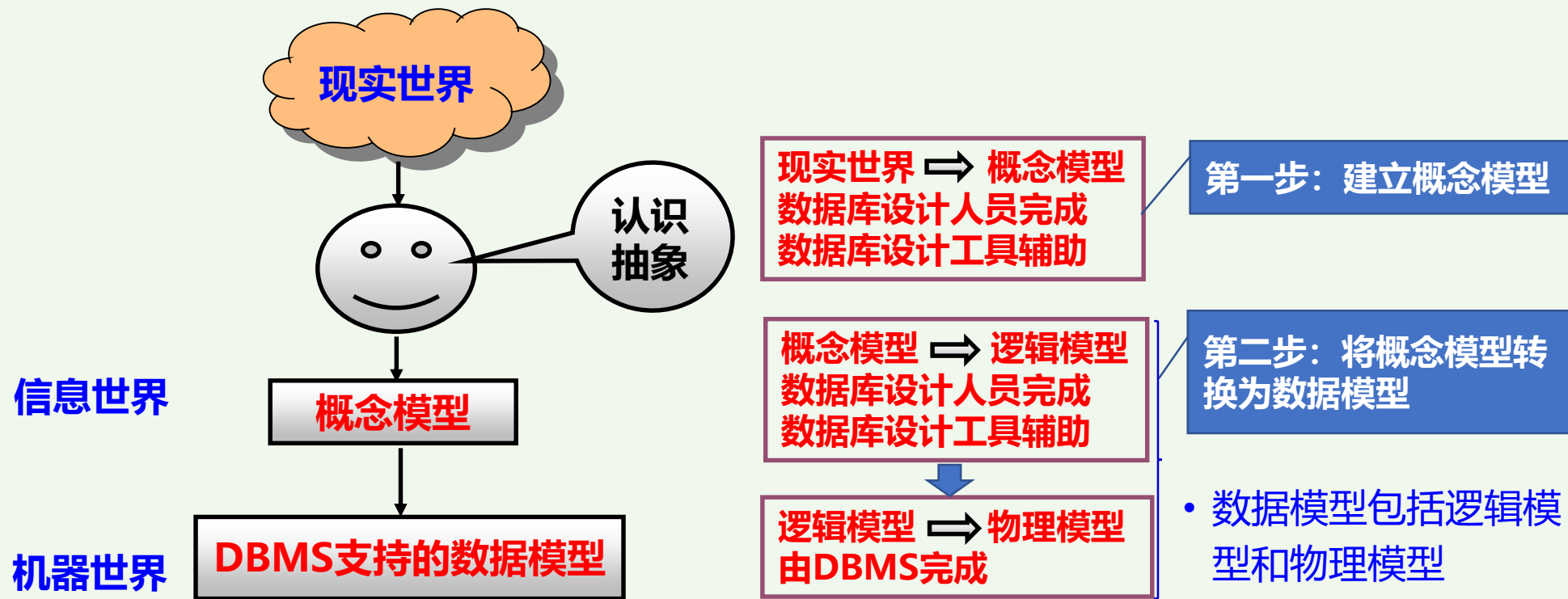


Fig.数据建模的两步抽象过程

- 概念模型是按用户观点来对数据建模，主要用于数据库设计
- 概念模型的用途
 - 是现实世界到机器世界的一个中间层次
 - 用于信息世界的建模
 - 现实世界到信息世界的第一层抽象
 - 数据库设计的有力工具
 - 数据库设计人员和用户之间进行交流的语言
- 对概念模型的基本要求
 - 较强的语义表达能力
 - 简单、清晰、易于用户理解



■ 实体(Entity)

- 客观存在并可相互区别的**事物**称为实体，可以是具体的人、事、物、抽象的概念或联系
- 例：一个学生，某个订单，论文指导关系

■ 属性(Attribute)

- 实体所具有的某一**特性**称为属性，一个实体可由若干个属性来刻画
- 例：学生实体包括属性：学号，姓名，性别，出生日期，主修专业等

■ 码/键(Key)

- **唯一标识实体**的属性集称为码
- 例：学号是学生实体的码



■ 实体型(Entity type)

- 用**实体名及其属性名集合**来抽象和刻画同类实体称为实体型
- 例：学生(学号，姓名，性别，出生日期，主修专业)就是一个实体型

■ 实体集(Entity set)

- 同一类型实体的集合称为实体集。例：全体学生就是一个实体集

■ 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(型)内部的联系和实体(型)之间的联系
- **实体内部的联系**通常是指组成实体的各属性之间的联系
- **实体之间的联系**通常是指不同实体集之间的联系
- 实体之间的联系有一**对一**、**一对多**和**多对多**等多种类型



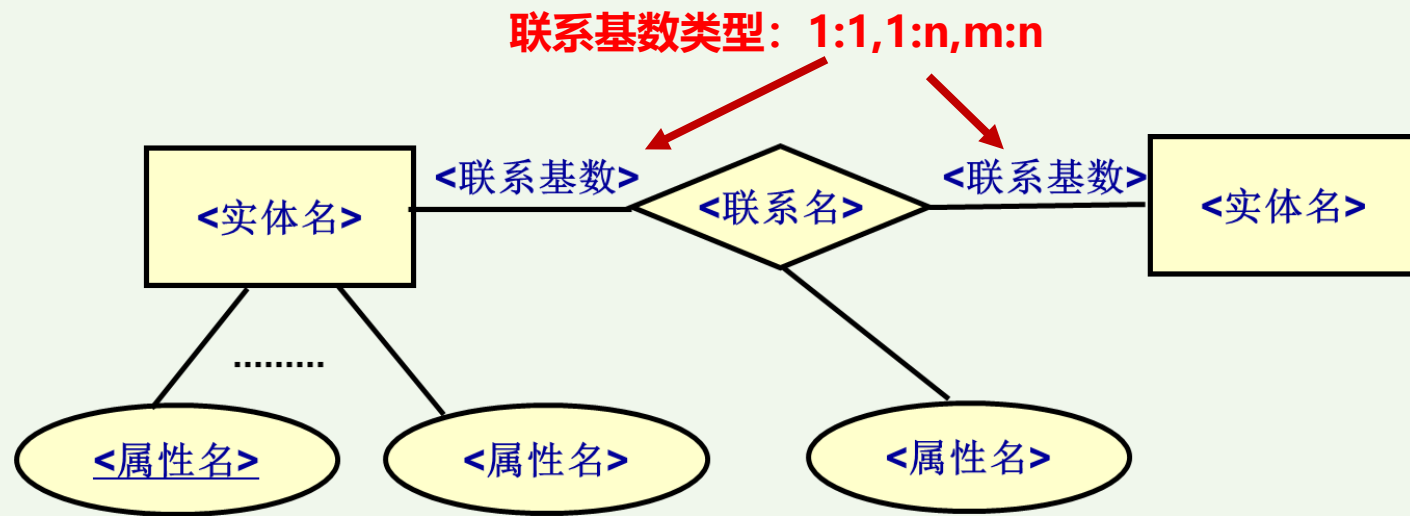
- 实体-联系图(Entity-relationship Diagram, ERD)
 - 概念建模的一种常用表示方法/工具
 - 也称为E-R方法(approach)/E-R图(diagram)/E-R模型(model)
 - 由P.P.S.Chen(陈品山)1976年提出

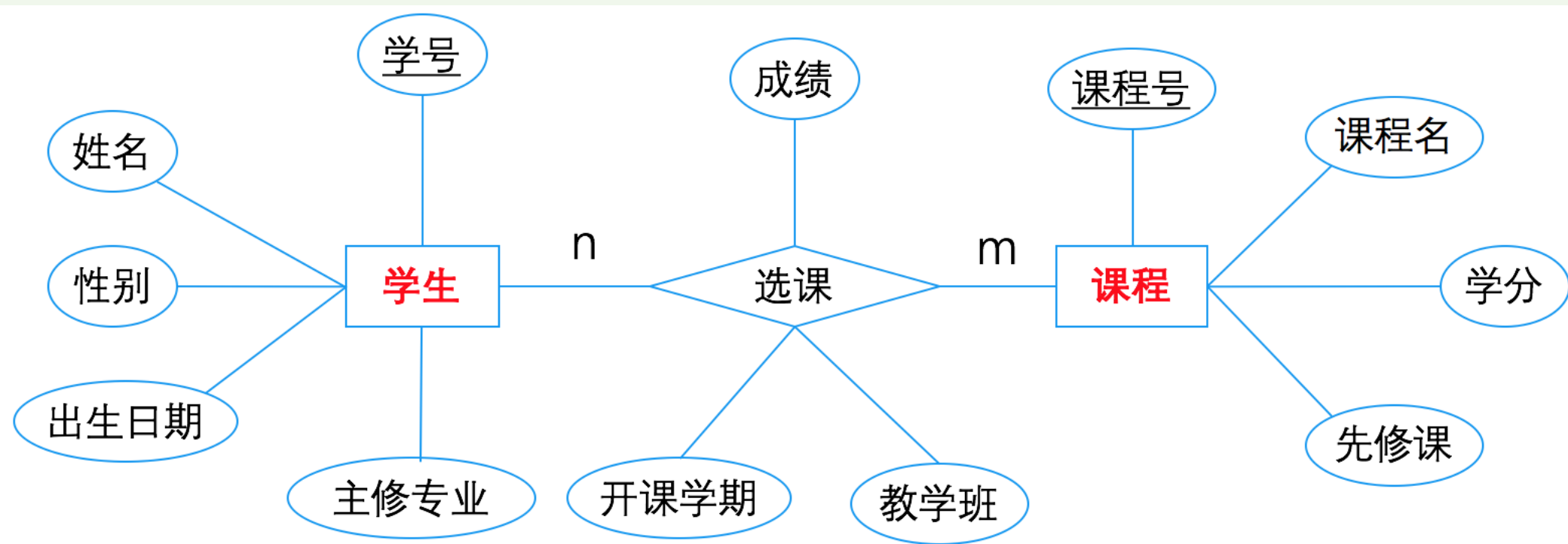


Peter Chen. The Entity-Relationship Model--Toward a Unified View of Data. ACM Transactions on Database Systems, Vol. 1(1), p.9-36,1976

- ERD表示要点

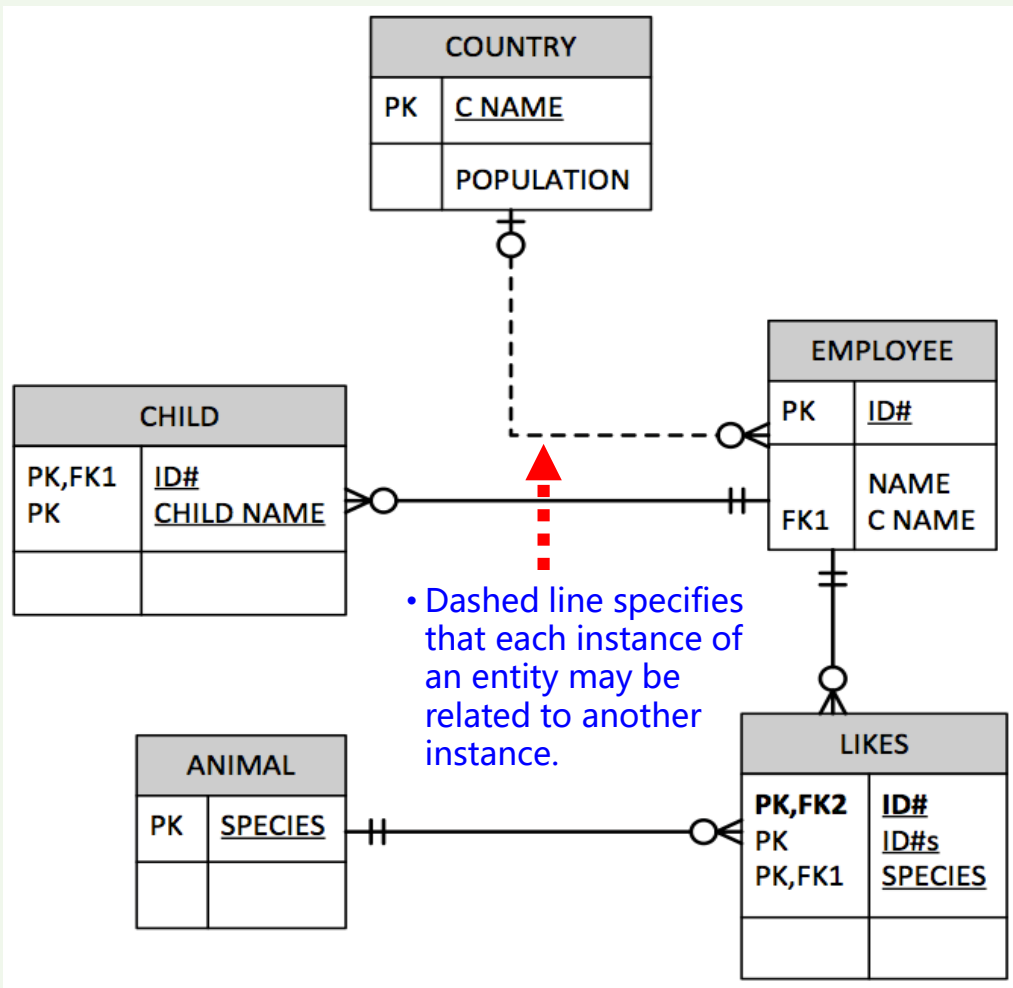
- 陈氏法





- 抽象了学校中的学生和课程两个客观事物：学生实体和课程实体
- 抽象了现实世界中事物之间的联系：
 - 一门课程可以有多名学生选修，一个学生可以选修多门课程
 - 用课程实体与学生实体多对多(m:n)联系来描述





Symbol	Meaning
	One—Mandatory
	Many—Mandatory
	One—Optional
	Many—Optional

<https://www.conceptdraw.com/examples/crowfoot-notation>



在教务系统中有业务规则(**business rule**)如下:

- 每个学生都有一个座位, 每个座位只能允许至多一个学生坐;
- 每个老师可以教多门课程, 也可以不教课, 每门课只能由一个老师教;
- 每个学生可以选修多门课, 也可以不选任何课, 一门课可以被多个学生选, 也可以允许没有学生选。
- 学生, 座位, 教师和课程实体名分别为: **student, seat, lecturer, course**

请根据以上语义分别使用陈氏记法和鱼尾纹记法画出该**E-R图**



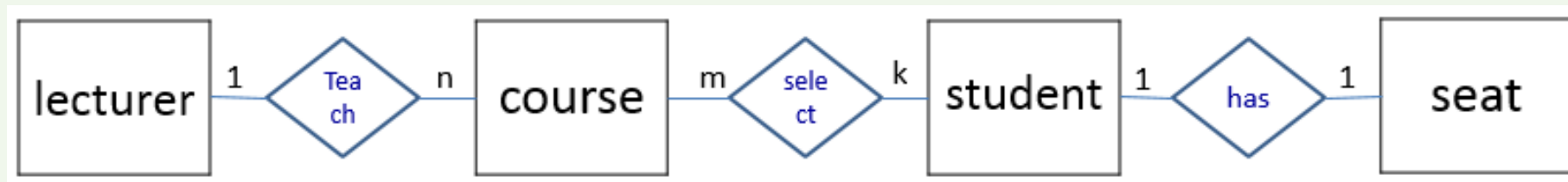


Fig 1. Chen's Notation

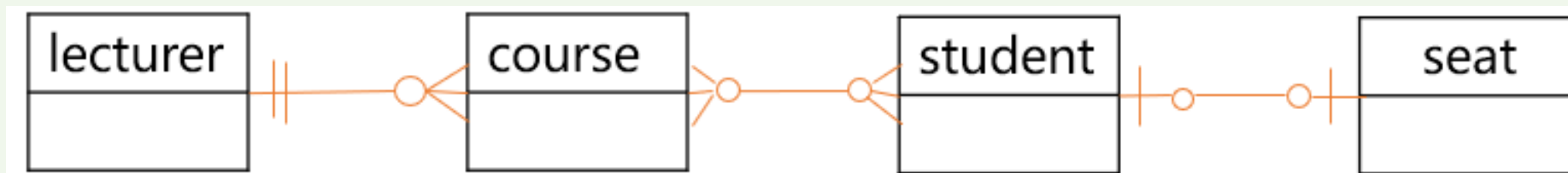
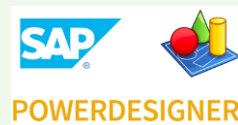


Fig 2. Crow's Foot Notation

- 手工
- 数据建模工具

- SAP Sybase PowerDesigner

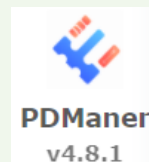


- Erwin 

- Oracle SQL Developer Data Modeler

- Microsoft Visio

- 国产开源免费数据建模工具 **PDManer**



- <https://gitee.com/robergroup/pdmaner/releases>

- ...



■ 数据结构、数据操作、完整性约束是数据模型的三要素

数据结构	数据操纵	完整性约束
<ul style="list-style-type: none">• 描述数据库的组成对象以及对象之间的联系• 描述的内容<ul style="list-style-type: none">– 与对象的类型、内容、性质有关– 与数据间联系有关的对象• 是刻画一个数据模型性质的最重要方面• 是对系统静态特性的描述	<ul style="list-style-type: none">• 是指对数据库中对象(型)的实例(值)允许执行的操作集合, 包括操作及有关的操作规则。• 数据库主要操作<ul style="list-style-type: none">– 查询– 更新(插入、删除、修改)• 数据模型必须定义<ul style="list-style-type: none">– 操作的确切含义– 操作符号– 操作规则(如优先级)– 实现操作的语言• 是对系统动态特性的描述	<ul style="list-style-type: none">• 是一组完整性规则• 完整性规则<ul style="list-style-type: none">– 是给定的数据模型中数据及其联系所具有的制约和依存规则– 用以限定符合数据模型的数据库状态以及状态的变化, 以保证数据的正确、有效和相容• 数据模型应该反映和规定其必须遵守的基本和通用的完整性约束。• 数据模型都是逻辑上的• 物理模型<ul style="list-style-type: none">– 数据模型以一定的组织方式存储在DBMS中, 是数据模型在DBMS内部的物理存储结构。– 即, 存储和存取



■ 四种数据模型的比较

	层次模型	网状模型	关系模型	面向对象模型
创始	1968 IBM IMS系统	1969 CODASYL DBTG报告	1970 E.F.Codd提出	20世纪80年代
数据结构	复杂（树结构）	复杂（有向图结构）	简单（二维表）	复杂（嵌套，递归）
数据联系	通过指针	通过指针	通过表间公共属性	面向对象标识
查询语言	过程性语言	过程性语言	非过程性语言	面向对象语言
典型产品	IMS	IDS/II, IMAGE/3000, IDMS	Oracle, openGauss	ONTOS DB
盛行期	20世纪70年代	20世纪70年代到80年代中期	20世纪80年代到现在	20世纪90年代初

• 非关系数据模型请自行看书，课堂不讲



- 关系模型是关系数据模型的简称
- 关系模型是最重要的一种数据模型
- 关系数据库系统采用关系模型作为数据的组织方式
- 由IBM San Jose的Edgar.F.Codd于1970年首次提出
 - Codd也因此项杰出工作获得1981年的图灵奖
- 1980年代以来，计算机厂商新推出的数据库管理系统几乎都支持关系模型



学生登记表数据结构

学号 Sno	姓名 Sname	性别 Ssex	出生日期 Sbirthdate	主修专业 Smajor
20180001	李勇	男	2000-3-8	信息安全
20180002	刘晨	女	1999-9-1	计算机科学与技术
20180003	王敏	女	2001-8-1	计算机科学与技术
20180004	张立	男	2000-1-8	计算机科学与技术
20180005	陈新奇	男	2001-11-1	信息管理与信息系统
20180006	赵明	男	2000-6-12	数据科学与大数据技术
20180007	王佳佳	女	2001-12-7	数据科学与大数据技术

- 从**用户观点**下，关系模型是由一组关系组成
- 每个关系的数据结构是一张规范的**二维表**

- **关系(relation)**
 - 一个关系对应一张二维表
- **元组(tuple)**
 - 表中的一行即为一个元组
- **属性(attribute)**
 - 表中的一列即为一个属性，每列的名称即为属性名
- **码(key)**
 - 又称码或键。表中的某一个属性或一组属性，它的值可以唯一确定一个元组
- **域(domain)**
 - 是一组具有相同数据类型的值的集合。属性的取值范围来自某个域
- **分量(component)**
 - 元组中的一个属性值
- **关系模式**
 - 对关系的描述
 - 表示格式：关系名(属性1,属性2,...,属性n)



- 关系模型要求关系必须是**规范化 (normalization)**的
 - 关系必须满足一定的规范条件
 - 最基本的规范条件**: 关系的每一个分量必须是一个不可分的数据项
 - 即, 不允许表中套表**

学号	姓名	性别	出生日期	主修专业	联系方式		
					手机号	Email	微信号
20180001	李勇	男	2000-3-8	信息安全	18301200745	liyong@ qq. com	liyong@ ruc
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

联系方式属性是可分的数据项,不符合关系模型要求, 故上表不是一个关系模型

表1.4 关系术语与现实生活中表格使用的术语对比

关系术语	一般表格的术语
关系名	表名
关系模式	表头 (表格的描述)
关系	(一张)二维表
元组	记录或行
属性	列
属性名	列名
属性值	列值
分量	一条记录中的一个列值
非规范关系	表中有表(大表中嵌有小表)

- 关系模型中的数据操纵是集合操作，操作对象和操作结果都是关系
 - 查询
 - 插入
 - 删除
 - 更新
- 存取路径对用户隐蔽，用户只要指出“干什么”，不必详细说明“怎么干”
- 关系的完整性约束包括如下三大类：
 - 实体完整性
 - 参照完整性
 - 用户定义的完整性



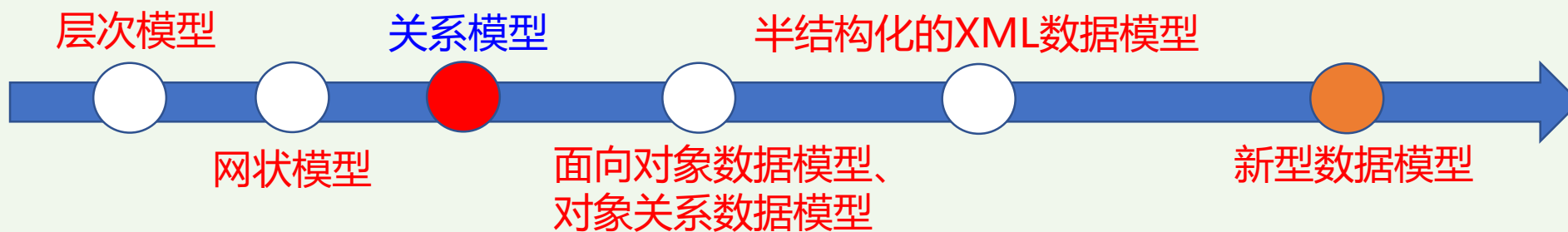
■ 优点

- 建立在严格的数学概念的基础上
- 概念单一
 - 实体和实体之间联系都用关系来表示
 - 对数据的检索和更新结果也是关系
- 存取路径对用户隐蔽
 - 具有更高的数据独立性，更好的安全保密性
 - 简化了程序员的工作和数据库开发建立的工作

■ 缺点

- 存取路径对用户隐蔽，查询效率往往不如层次模型和网状模型
- 为提高性能，必须对用户的查询请求进行优化，增加了开发DBMS的难度





- 相比于关系模型，这些模型按照数据模型应该具备的三个基本要素来衡量**不严格**，例如对这些数据的完整性约束就需要研究和发展

- 键值对数据模型 (Key-Value)
- 文档数据模型
- 图数据模型
- 时序数据模型
- 时空数据模型
- 流数据模型
- 多媒体数据模型
- 向量数据模型

- 数据模型是对现实世界客观对象的抽象
- 数据模型应该满足三方面要求：
 - 比较真实地模拟世界
 - 容易为人所理解
 - 便于在计算机上实现
- 数据建模的过程划分为两个层次
 - 概念模型：按照用户的观点对数据和信息建模
 - 数据模型：按计算机系统的观点对数据建模
 - 包括逻辑模型和物理模型



- 从数据库管理系统角度看，数据库系统通常采用三级模式结构，是数据库系统内部的体系结构
- 根据计算机的系统结构，从数据库最终用户角度看，数据库系统的外部体系结构分为：
 - 集中式结构
 - 客户机/服务器（浏览器/应用服务器/数据库服务器）
 - 并行结构
 - 分布式结构
 - 云结构
- 数据库系统的内部体系结构与外部体系结构之间的关系见第7部分



■ “型” 和 “值” 的概念

– 型(type)

- 类型的简称
- 是对一类数据的结构和属性的说明
- 例子：学生记录：(学号，姓名，性别，出生日期，主修专业)

– 值(value)

- 是型的一个具体赋值
- 例子：上面型的一个实例：(20180003, 王敏, 女, 2001-8-1, 计算机科学与技术)



■ 模式(Schema)

- 数据库中全体数据的逻辑结构和特征的描述
- 是型的描述, 不涉及具体值
- 反映的是数据的结构及其联系
- 模式是相对稳定的
 - 模式依赖于系统的业务逻辑

■ 实例(Instance)

- 模式的一个具体值
- 反映数据库某一时刻的状态
- 同一个模式可以有很多实例
- 实例随数据库中数据的更新而变动

选课模式:
选课(学号, 课程号, 成绩)



学号	课程号	成绩
2019001	DB001	87
2019007	OS239	73
...

2022学年第一学期选课信息



学号	课程号	成绩
2020001	DB001	62
2020007	OS239	90
...

2023学年第一学期选课信息



两个实例



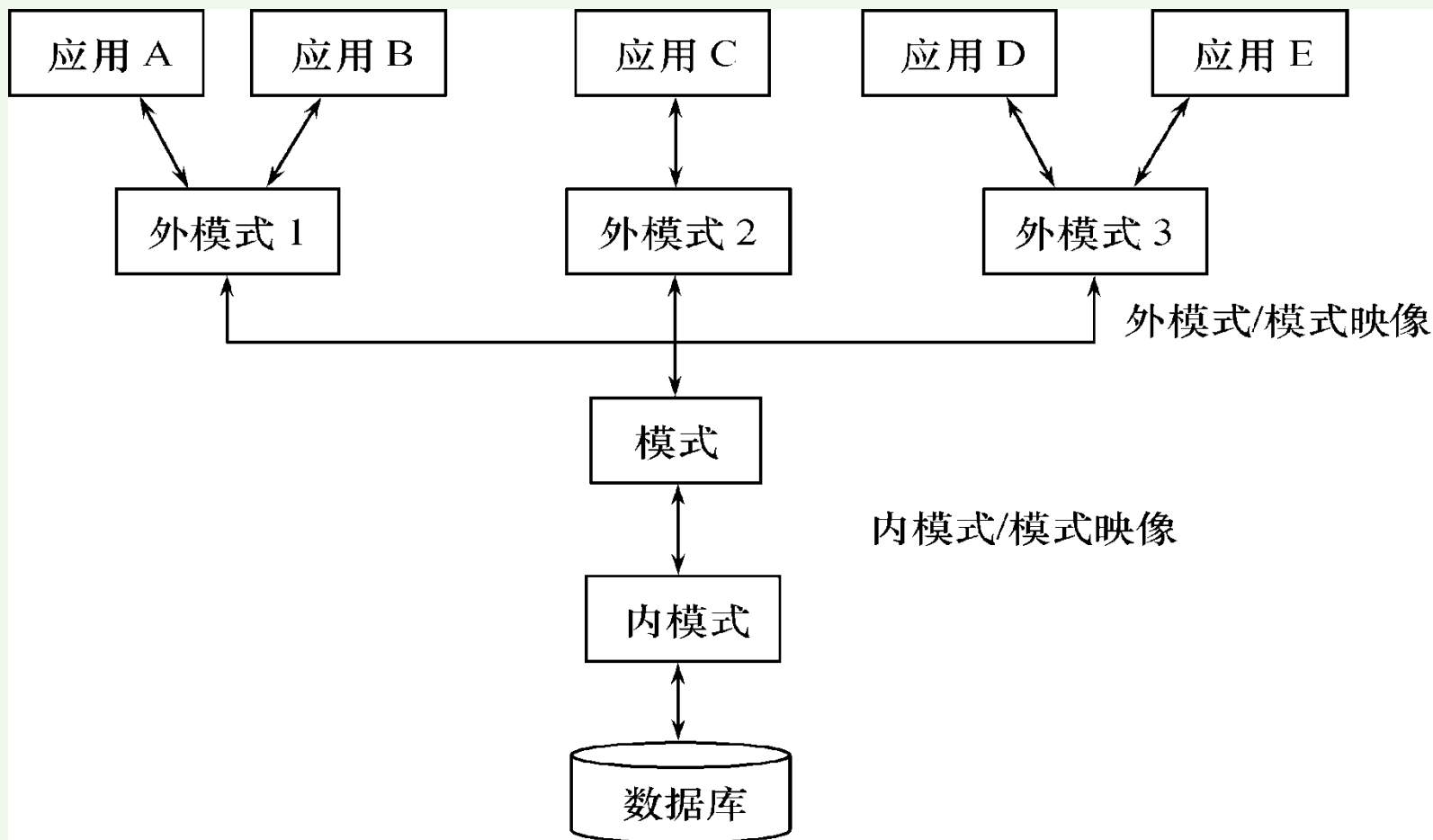


Fig. ANSI-SPARC Architecture

(American National Standards Institute, Standards Planning and Requirements Committee)

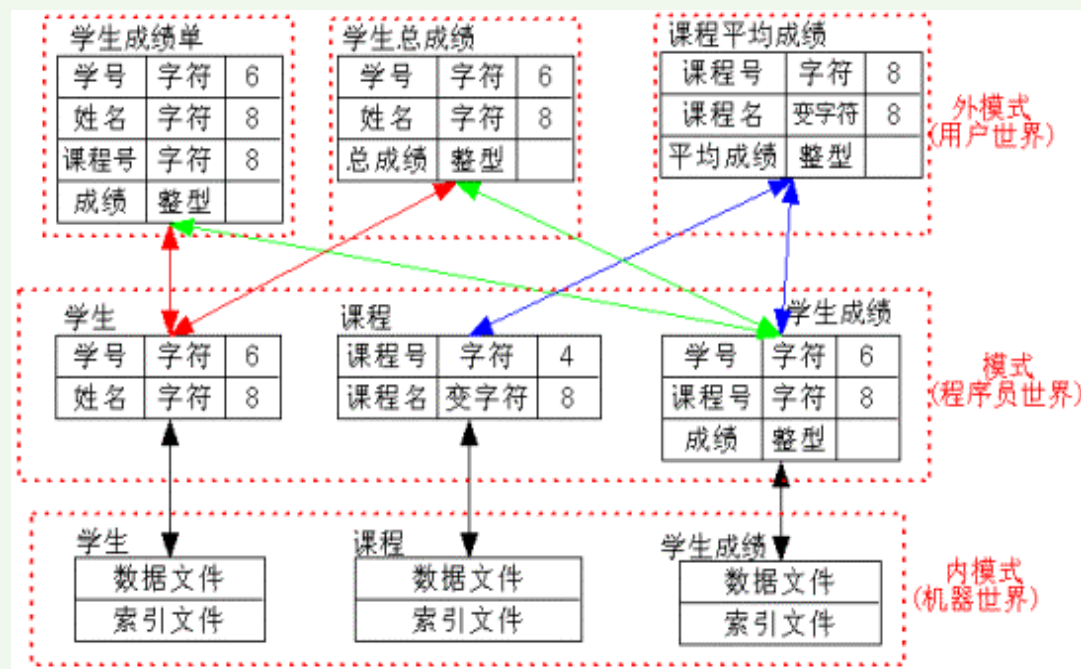
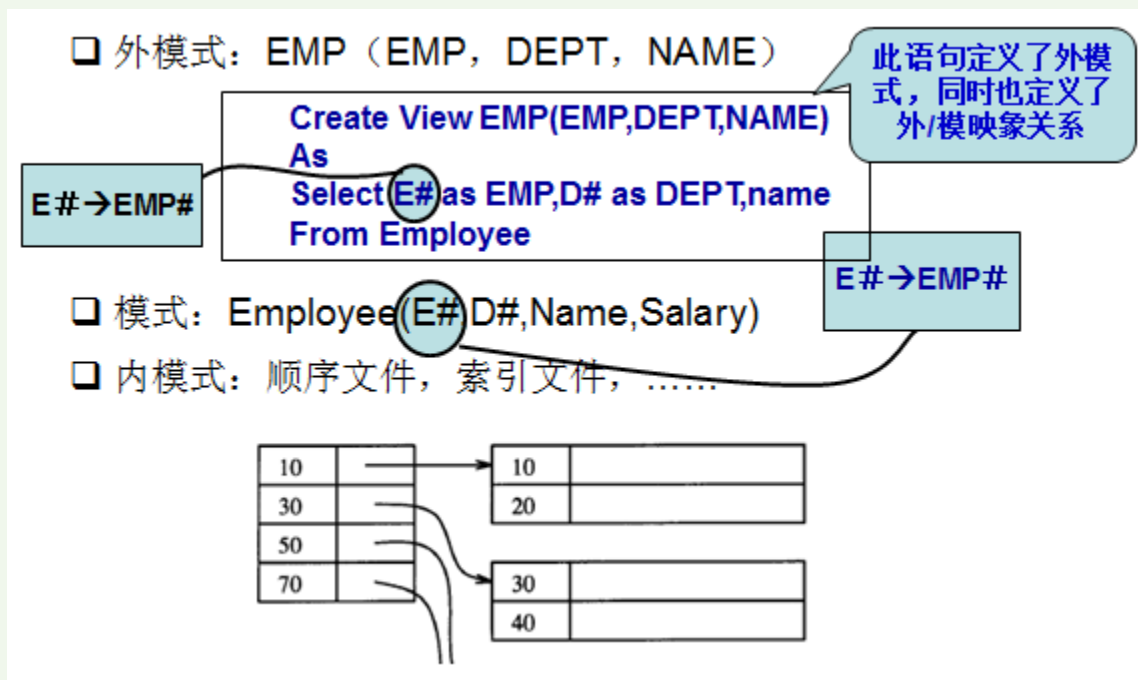


	外模式 (External Schema)	模式 (Schema)	内模式 (Internal Schema)
定义	也称子模式或用户模式 还称用户级模式	也称逻辑模式 还称概念级模式	也称存储模式 还称物理级模式
	是数据库用户能够看见和使用的局部数据的逻辑结构和特征的描述	是数据库中全体数据的逻辑结构和特征的描述, 它包括: 数据的逻辑结构、数据之间的联系和与数据有关的安全性、完整性要求。	它是数据物理结构和存储方式的描述
特点1	是各个具体用户所看到的数据视图, 是用户与DB的接口。	是所有用户的公共数据视图	数据在数据库内部的表示方式
特点2	可以有多个外模式	只有一个模式	只有一个内模式
特点3	针对不同用户, 有不同的外模式描述。每个用户只能看见和访问所对应的外模式中的数据, 数据库中其余数据是不可见的。所以外模式是保证数据库安全性的一个有力措施。	数据库模式以某一种数据模型(层状、网状、关系)为基础, 统一综合地考虑所有用户的需求, 并将这些需求有机地结合成一个逻辑整体。	
特点4	面向应用程序或最终用户	由DBA定义	以前由DBA定义, 现基本由DBMS定义



- **三级模式**是对数据的三个抽象级别
- **两级映象**在**DBMS**内部实现三个抽象层次的**联系和转换**
 - 外模式/模式映象
 - 模式/内模式映像
- **外模式/模式映象**
 - 当模式改变时，数据库管理员修改有关的外模式/模式映象，使外模式保持不变。应用程序是依据数据的外模式编写的，外模式不变，应用程序就没必要修改。所以外模式/模式映像功能保证了数据与程序的逻辑独立性，简称**数据的逻辑独立性**
- **模式/内模式映象**
 - 当数据库的存储结构改变了，模式/内模式映像作相应的改变，以使模式保持不变。模式不变，与模式没有直接联系的应用程序也不会改变，所以模式/内模式映像功能保证了数据与程序的物理独立性，简称**数据的物理独立性**





- 模式/内模式映象定义了数据全局逻辑结构与存储结构之间的对应关系
 - 如说明逻辑记录和字段在内部是如何表示的
- 数据库中模式/内模式映象是唯一的
- 该映象定义通常包含在模式描述中

- 每一个外模式, 数据库系统都有一个外模式/模式映象定义外模式与模式之间的对应关系
- 映象定义通常包含在各自外模式的描述中

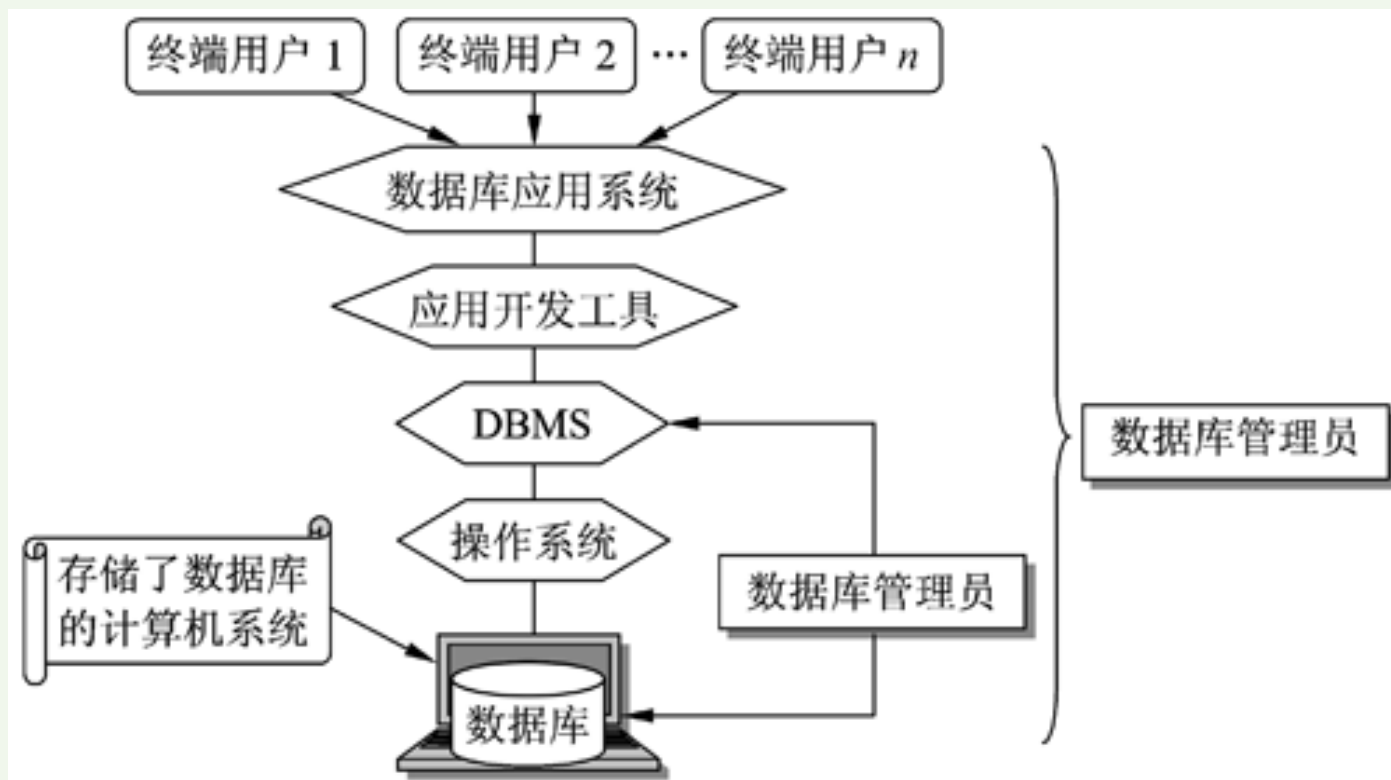


5

5.4数据库三级模式结构小结

模式	内模式	外模式	特定应用程序	二级映像
<ul style="list-style-type: none">• 即全局逻辑结构• 是数据库的核心与关键• 独立于数据库的其他层次• 设计数据库模式结构时应首先确定数据库的逻辑模式	<ul style="list-style-type: none">• 依赖于它的全局逻辑结构• 独立于数据库的用户视图，即外模式• 独立于具体的存储设备• 将全局逻辑结构中所定义的数据结构及其联系按照一定的物理存储策略进行组织，以达到较好的时空效率	<ul style="list-style-type: none">• 面向具体的应用程序• 定义在逻辑模式之上• 独立于存储模式和存储设备• 当应用需求发生较大变化，相应外模式不能满足其视图要求时，该外模式就得做相应改动• 设计外模式时应充分考虑到应用的扩充性	<ul style="list-style-type: none">• 在外模式描述的数据结构上编制的• 依赖于特定的外模式• 与数据库的模式和存储结构独立• 不同的应用程序有时可以共用同一个外模式	<ul style="list-style-type: none">• 保证了数据库外模式的稳定性• 从底层保证了应用程序的稳定性，除非应用需求本身发生变化，否则应用程序一般不需要修改• 数据与程序之间的独立性使得数据的定义和描述可以从应用程序中分离出去• 数据的存取由DBMS管理，简化了APPs的编制和开发与维护成本





数据库系统 = 硬件 + 数据库 + OS + DBMS + 应用系统 + 应用开发工具 + DBA + 用户

软件平台

人员



■ 数据库管理员(DBA)

- 决定数据库中的信息内容和结构、存储结构和存取策略；定义数据的安全性要求和完整性约束条件；监控数据库的使用和运行；数据库的改进和重组重构

■ 系统分析员

- 负责应用系统的需求分析和规范说明，与用户和**DBA**一起确定系统的软硬件配置，并参与系统的总体设计

■ 数据库设计人员

- 负责数据库中数据的确定和数据库各级模式的设计，常由**DBA**担任

■ 应用程序员

- 以外模式为基础进行应用系统的设计、编码、测试和安装

■ 最终用户

- 具体操作应用系统，通过应用系统客户端的用户界面使用数据库以完成业务活动



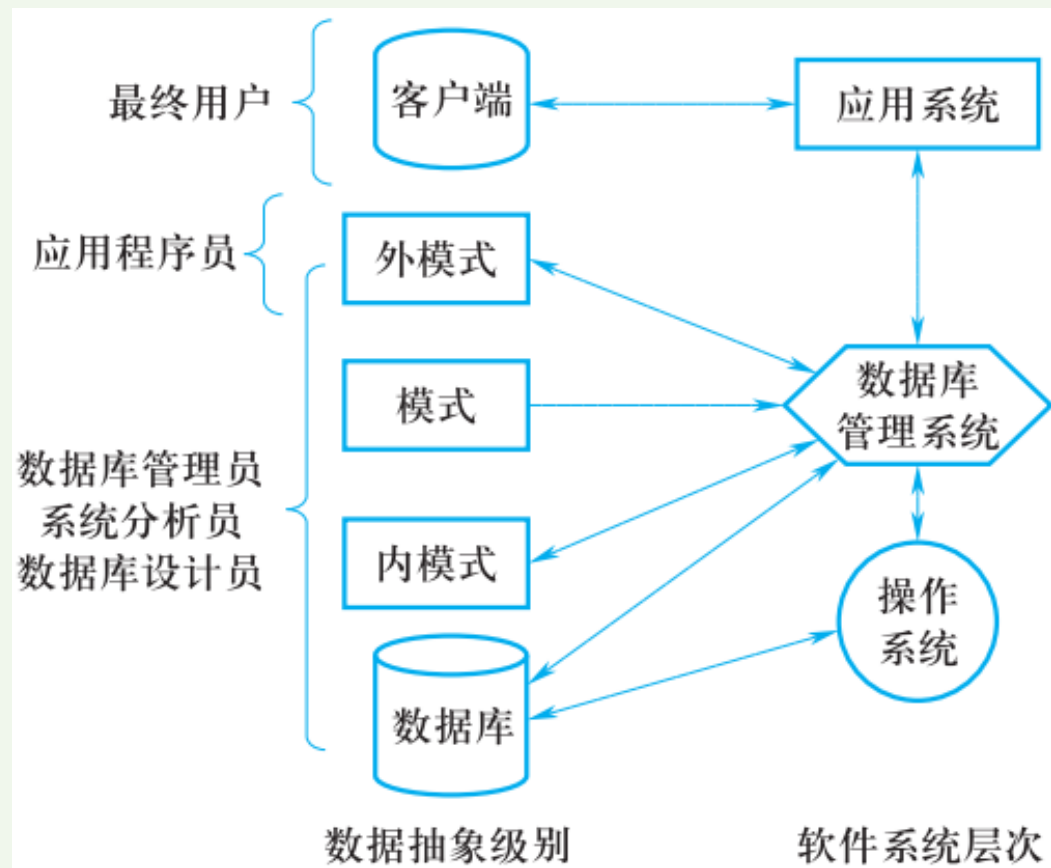
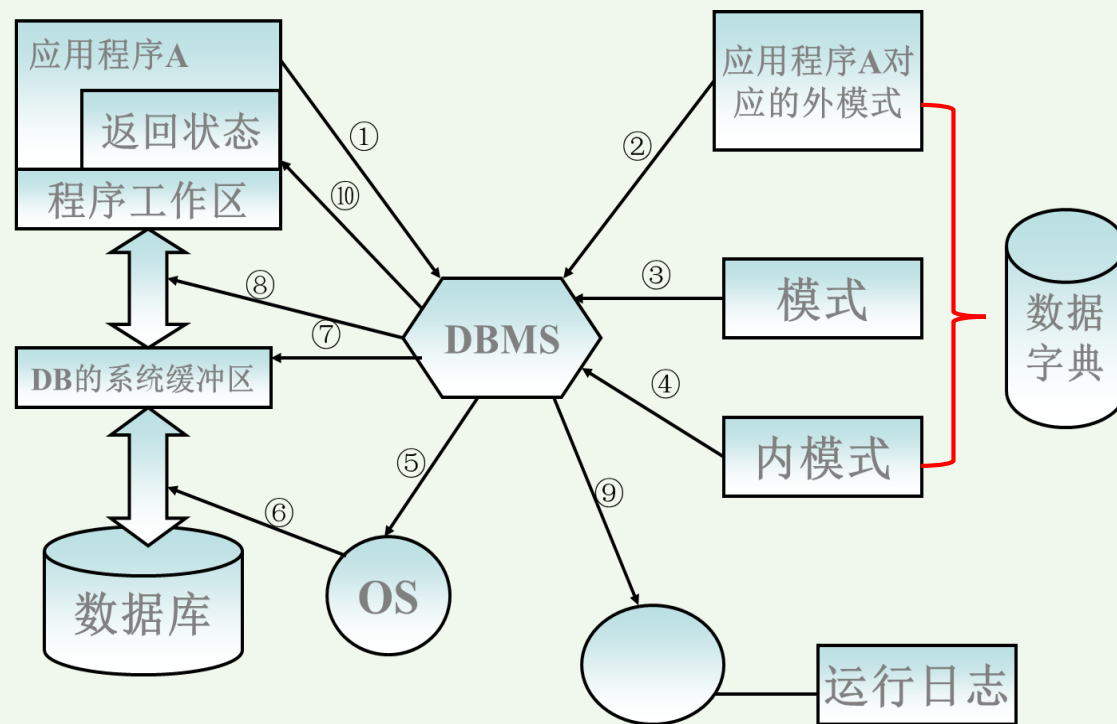


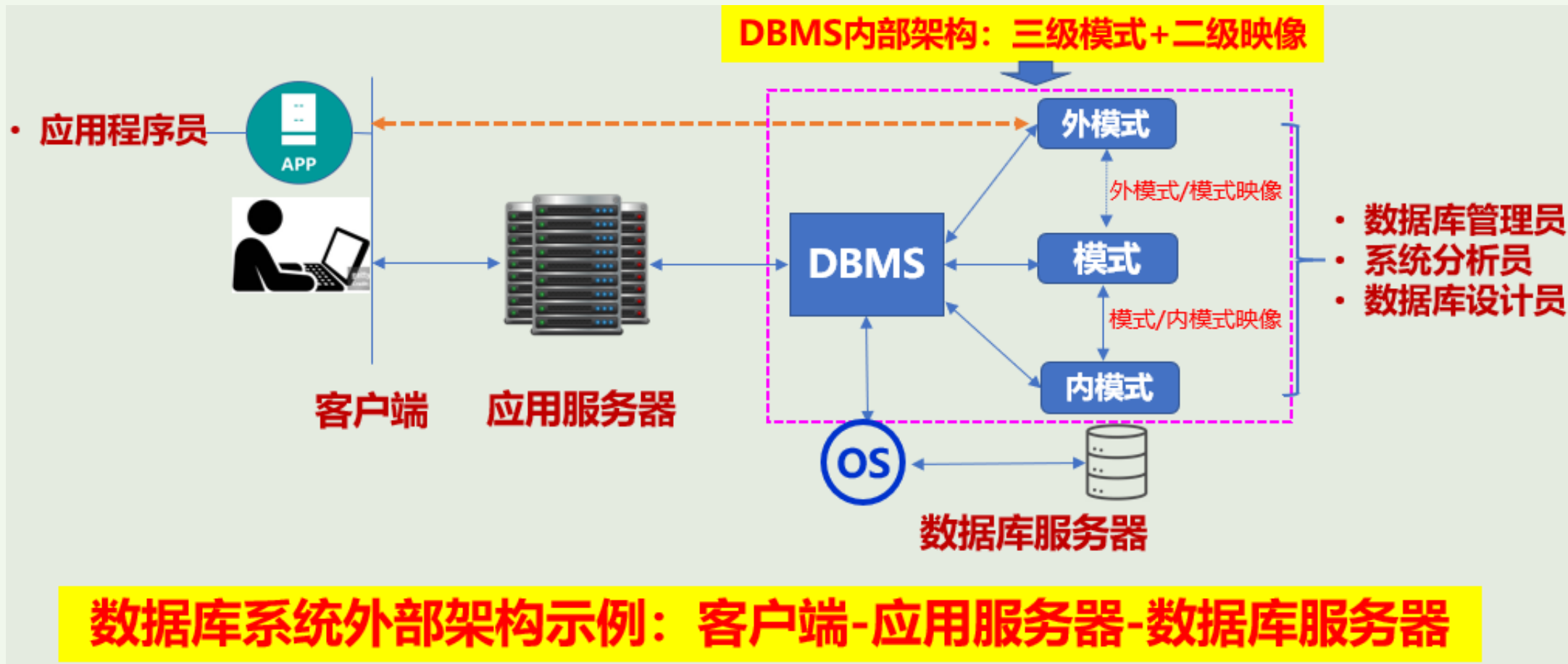
图1.16 数据库系统各种人员的数据视图



用户访问数据库的过程示意图



集中式DBS	客户机-服务器DBS	并行DBS	分布式DBS	云DBS
<ul style="list-style-type: none"> 数据库和应用程序都在同一台计算机上 小型机和大型机上的集中式DBS一般是多用户系统，多个用户通过各自的终端运行不同的应用系统，共享DB 微型计算机上的集中式DBS一般是单用户的 	<ul style="list-style-type: none"> DBMS、DB驻留在服务器上 应用程序放置在客户机上(微型计算机或工作站) 客户机和服务器通过网络进行通信 客户机负责业务数据处理流程和应用程序的界面，当要存取数据库中的数据时就向服务器发出请求，服务器接收客户机请求后处理并将结果返回给客户机 三层结构：浏览器/应用服务器/数据库服务器 	<ul style="list-style-type: none"> 在并行计算机上运行的具有并行处理能力的DBS 是数据库技术与并行计算技术相结合的产物 并行计算机系统有共享内存型、共享磁盘型、非共享型及混合型 并行DBS发挥了多处理机的优势，采用并行查询处理技术和并行数据分布与管理技术，具有高性能、高可用性、高扩展性等优点 	<ul style="list-style-type: none"> 分布式DBS是指数据库中的数据在逻辑上是一个整体，但物理地分布着计算机网络上的不同节点上 网络中的每个节点独立处理本地数据库中的数据(称为场地自治)，执行局部应用 也可以执行全局应用，即通过网络通信系统同时存取和处理多个节点上数据库的数据 适用于企业部门分布的组织结构，降低费用，提高系统的可靠性和可用性，具有良好的可扩展性 	<ul style="list-style-type: none"> 云DBS把数据库部署或虚拟化在云计算环境下，通过计算机网络以服务的形式提供数据库功能，包括数据存储、数据更新、查询处理、事务管理 现在的云DBS是运行中集群上的并行DBS，能够较好地进行动态伸缩、按需分配计算资源和存储资源 缺点：云DBS存储的安全可靠、隐私保护等问题亟待研究解决



数据库系统概述	数据模型	DBS内部体系结构	DBS组成	DBS体系结构
<ul style="list-style-type: none">• 数据库的基本概念• 数据库技术产生和发展背景• 数据库系统的优点	<ul style="list-style-type: none">• 数据模型的三个要素• 概念建模：E-R模型• 三种数据模型	<ul style="list-style-type: none">• 三级模式• 两级映像	<ul style="list-style-type: none">• 硬件• 软件• 数据库• 应用程序• 开发工具• DBA• 人员	<ul style="list-style-type: none">• 集中式DBS• 客户-服务器DBS• 并行DBS• 分布式DBS• 云DBS



- 数据库系统的核心和基础是 ()
A.物理模型 B.概念模型 C.数据模型 D.逻辑模型
- 将现实世界抽象为信息世界的模型是 ()
A.物理模型 B.概念模型 C.关系模型 D.逻辑模型
- 能够保证数据库系统中的数据具有较高的逻辑独立性是 ()
A.外模式/模式映像 B.模式 C.模式/内模式映像 D.外模式
- **DBMS**是一类系统软件，它是建立在以下哪个系统之上的？ ()
A.应用系统 B.编译系统 C.操作系统 D.硬件系统
- 下列说法正确的是 ()
A.数据库的概念模型与具体的**DBMS**有关
B.三级模式中描述全体数据的逻辑结构和特征的是外模式
C.数据库管理员负责设计和编写应用系统的程序模块
D.从逻辑模型到物理模型的转换一般是由**DBMS**完成的



- 长期存储在计算机内，有组织的、可共享大量数据的集合是（ ）
A.数据 B.数据库 C.数据库管理系统 D.数据库系统
- 在数据库管理技术发展过程中，需要应用程序管理数据的是（ ）
A.人工管理阶段 B.人工管理阶段和文件系统阶段
C.文件系统阶段和数据库系统阶段 D.数据库系统阶段
- 在文件系统管理阶段，由文件系统提供数据存取方法，所以数据已经达到很强的独立性（对or错）
- 数据库管理系统是指在计算机系统中引入数据库后的系统，一般由DB, DBS, 应用系统和DBA组成(对or错)
- 在数据模型的组成要素中，数据结构是刻画一个数据模型性质最重要的方面，人们通常按照数据结构的类型来命名数据模型（对or错）
- 数据库系统的三级模式是对数据进行抽象的三个级别，把数据的具体组织留给DBMS管理（对or错）
- 层次模型是比网状模型更具普遍性的结构，网状模型是层次模型的一个特例（对or错）



- 数据库系统的逻辑模型按照计算机的观点对数据建模，主要包括____、____、____、面向对象模型、对象关系模式和半结构化数据模型等。
- 最常使用的概念模型是_____。
- 数据独立性是数据库领域的重要概念，包括数据的_____独立性和数据的_____独立性。
- 数据库的三级模式结构中，描述局部数据的逻辑结构和特征的是_____。
- 数据模型的组成要素中描述系统的静态特性和动态特性的分别是_____和_____。



- 教材第一章全部习题.
- 要求：作业在布置后一周内完成并提交到课程网站

