

第4章规模数据的有效管理



4.1数据管理基本概念



数据管理基本概念

数据

事物特性的反映和描述,是符号的集合。



张强 2018119030 男

李华 2018119060 女

数据是信息的符号表示(或称为载体)信息则是数据的内涵,是对数据语义的解释



数据 处理

对数据的收集、存储、加工、分类、检索、传播等一系列活动。

常规 管理 决策 支持 大数据 分析



提高管理效率



数据 库

的特点

将需要管理的数据按一定格式长期存储在计算机内的、有组织的、可共享的数据集合。

数据结构化

数据内部结构化,整体结构化,数据之间有联系。

数据库数据独立性

数据与程序独立,减少应用程序的维护和修改。

数据冗余低

只保留最少的数据重复(必要)

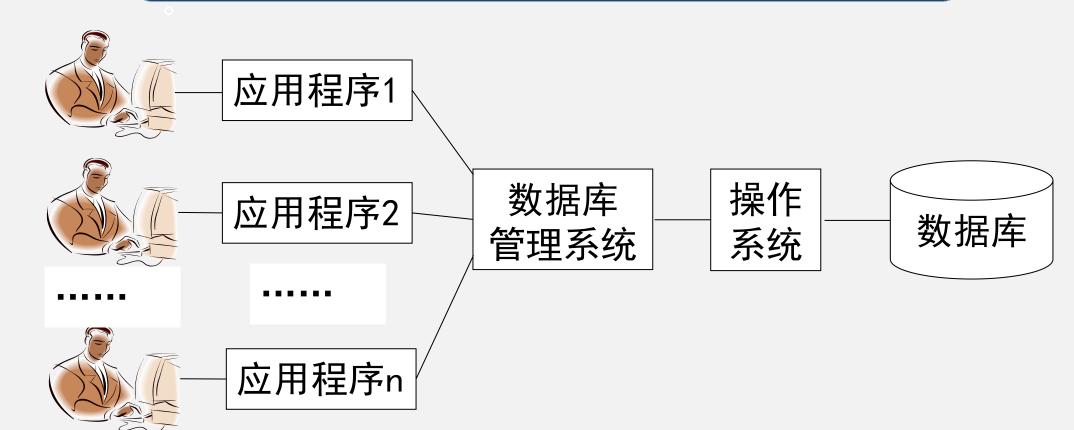
统一管理控制

数据库的管理由DBMS完成





位于用户与操作系统(OS)之间的一层数据管理 软件,为用户或应用程序提供访问数据库的方法







DDL,建立、修改数据库的库结构。包括模式定义、保密定义、完整性约束定义等。

DML,供用户实现对数据的追加、删除、 更新、查询等操作。

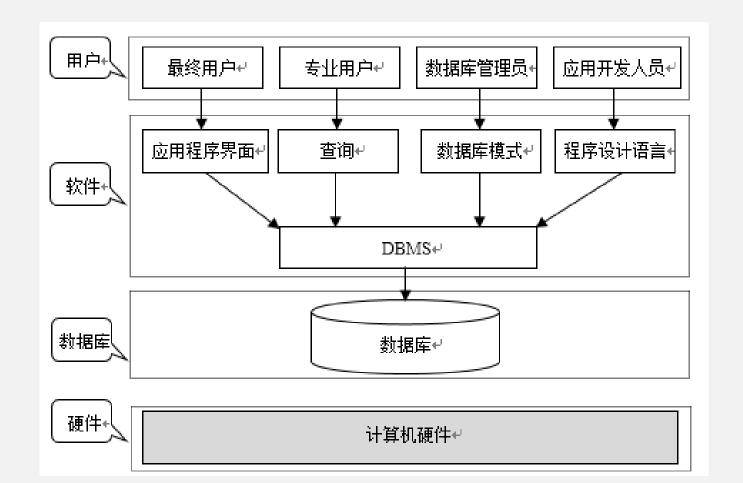
多用户环境下的并发控制、安全性检查和 存取限制控制、完整性检查和执行、运行 日志的组织管理、事务的管理和自动恢复。

分类组织、存储和管理各种数据,包括数据字典、用户数据、存取路径等,确定以何种文件结构和存取方式在存储级上组织这些数据。



数据 库系统

为适应管理的需要而开发建设的数据处理系统。



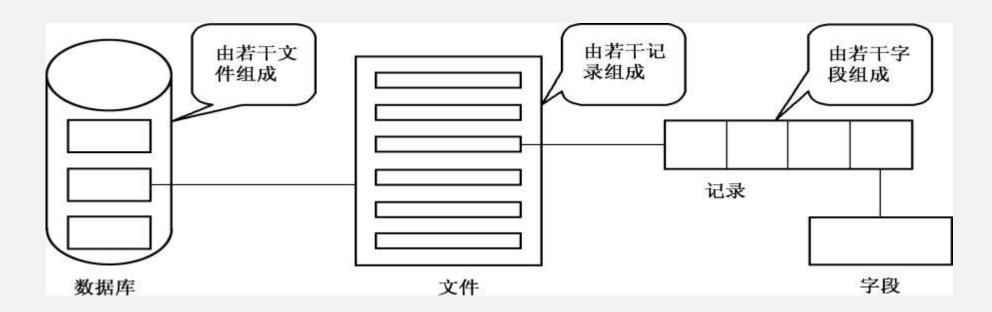


4.2数据的组织级别



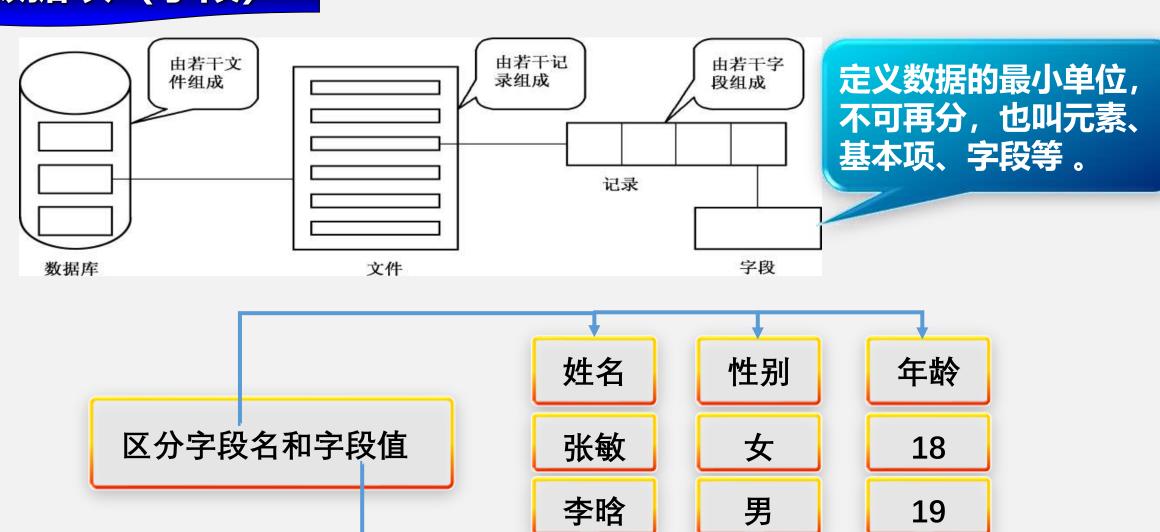
数据的组织级别

四级:数据项、记录、文件和数据库。



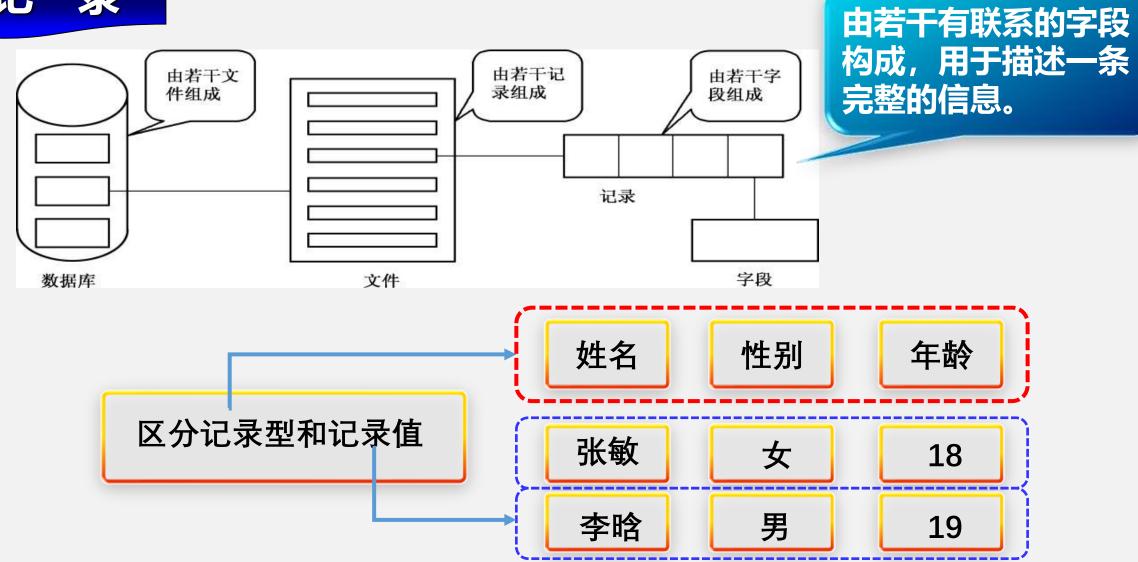


数据项 (字段)



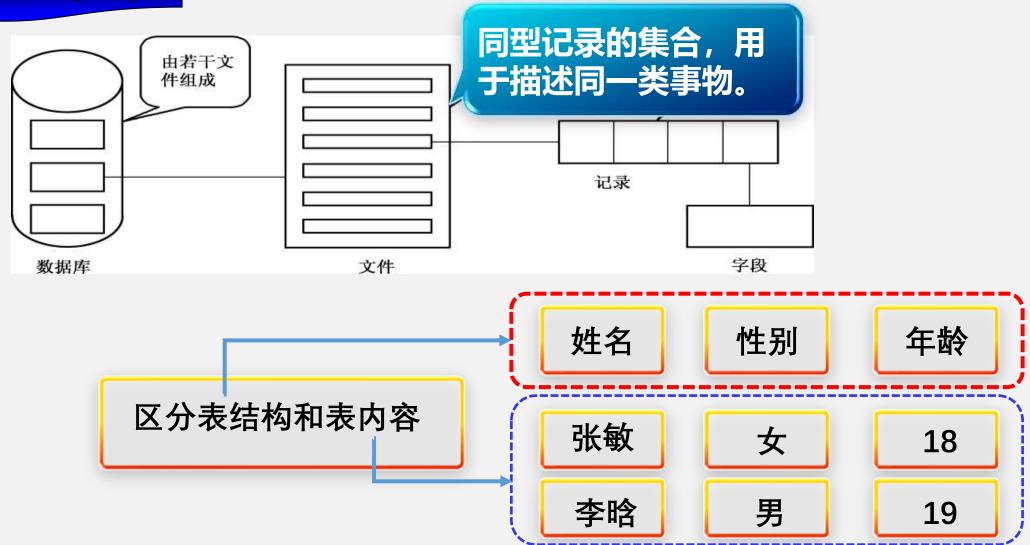


记录





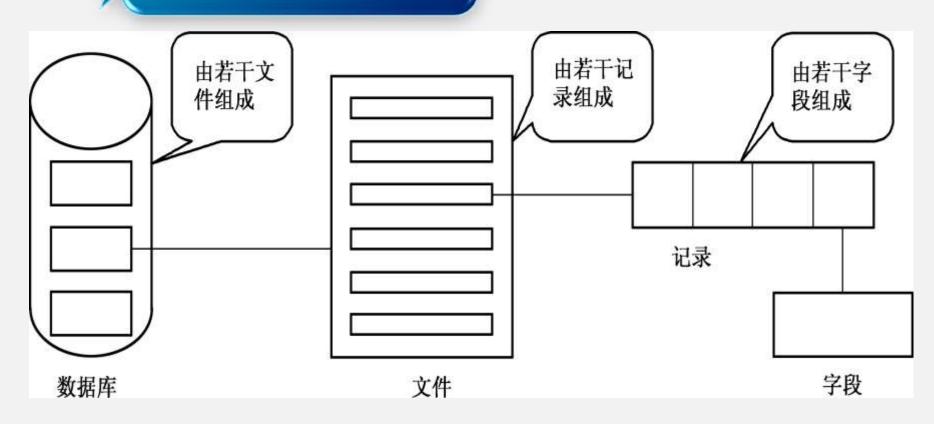
文件(表)



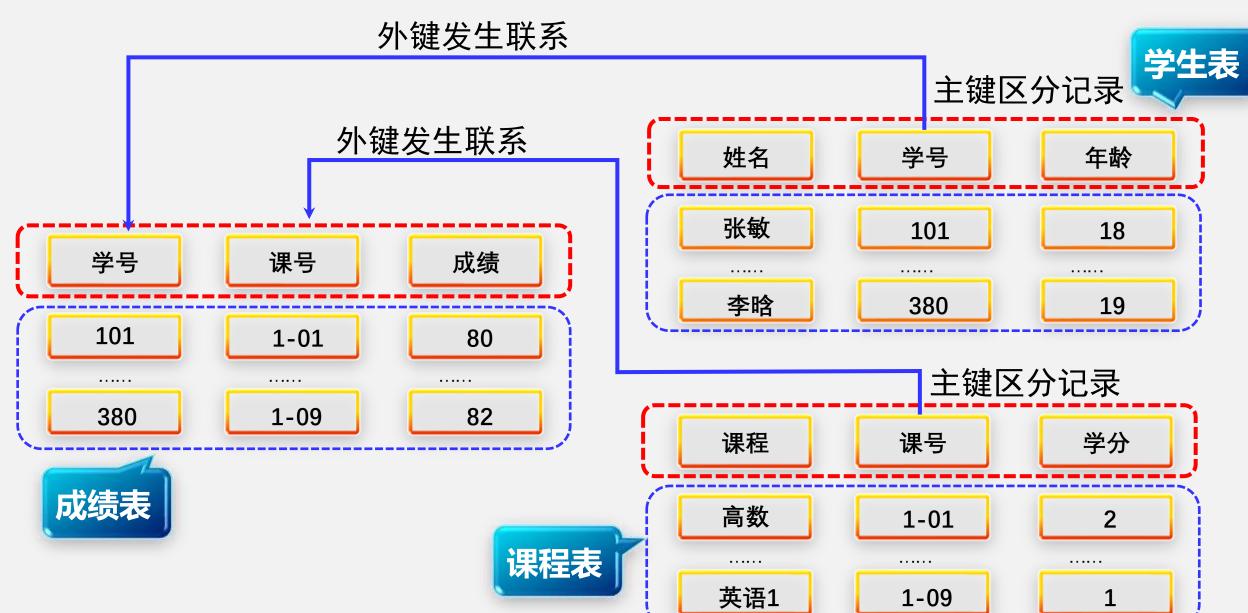


数据库

由若干张有联系的表组成,是管理的所有数据 的有组织的集合。







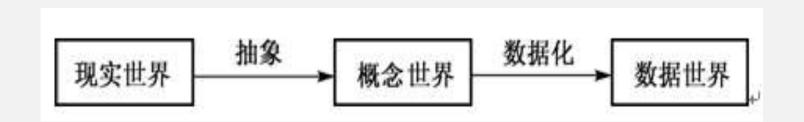


4.3 现实世界的数据表示



现实世界的数据表示

抽象表示三个阶段: 现实世界、概念世界、数据世界





基本概念

现实世界

现实世界里的客观事物是管理的对象,这些对象之间既有区别, 也有联系。

无穷事物,每个事物包含无穷特性

确定问题的边界: 哪些事物需要管理, 事物之间何种关系



概念世界

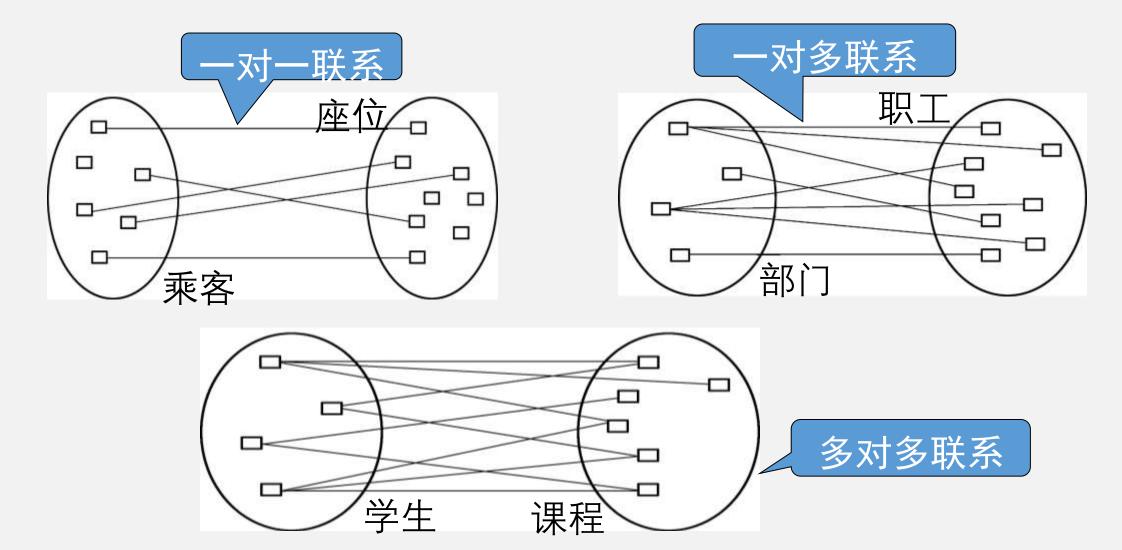
是现实世界在人脑中的反映,是对客观事物及其联系的抽象。也称为信息世界。

基本概念

- ◆ 实 体: 客观存在并可相互区别的事物
- ◆ 实体集: 同型实体的集合
- ◆ 实体型:用实体名及其属性名集合来描述同类实体
- ◆ 属 性:实体所具有的某一特性
- ◆ 域: 属性的取值范围
- ◆ 主 码: 唯一标识实体的属性集



联系:事物之间是有联系的,对应实体间的联系可以分为3类。





概念世界的表示

概念世界通过概念模型表示,与具体的DBMS 无关。

概念模型从用户的观点出发,将管理对象的客观事物及他们之间的联系,用容易理解的形式表述出来。

E-R图(实体联系图)是描述概念模型的主要工具。



矩形框:表示实体,框中记入实

学生

课程

体名。

菱形框:表示联系,框中记入联

学习

购买

系名。

椭圆:表示实体或联系的属性,属性名记

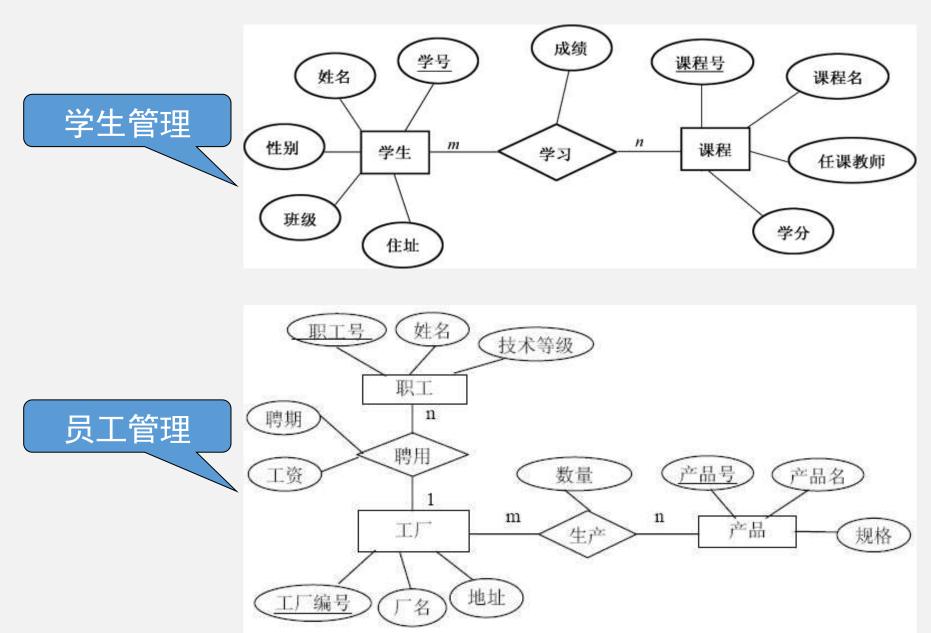
姓名

地址

入椭圆。

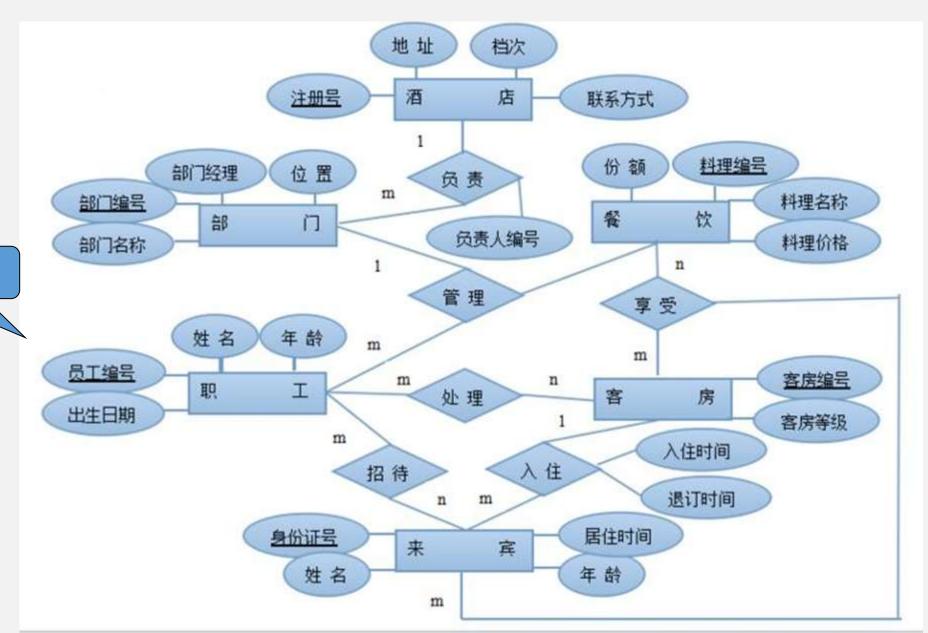
连线:用于实体与属性之间,实体与联系之间,联系与属性之间相连,并在直线上标注联系的类型。







餐饮管理







- (1) 确定所有的实体集合;
- (2) 选择实体应包含的属性;
- (3) 确定实体之间的联系;
- (4) 确定实体的关键字,用下划线表明关键字的属性组合;
- (5) 确定联系的类型,注明联系的类型(1:1, 1:n, m:n)。



数据世界

数据世界是信息世界进一步数据化的结果 。

基本概念

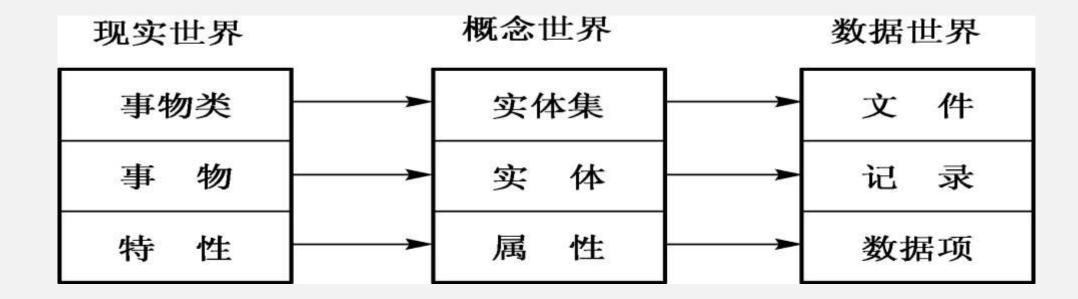
数据项:又称字段,是数据库数据中的最小逻辑单位,用来描述实体的属性。

记录:是数据项的集合,一个记录由若干个数据项组成,用来描述实体。

文件: 是一组同类记录的集合, 用来描述实体集。



对应关系





4.4 数据模型



数据模型

数据模型的概念

数据模型从抽象层次上描述了系统的静态特征、动态行为和约束条件,为数据库系统的信息表示与操作提供了一个形式框架。



数据模型的组成要素

三个要素

数据结构:主要描述数据的类型、内容、性质以及数据间的联系等。

数据操作:相应数据结构上的操作类型和操作方式。它是操作算符的集合,包括若干操作和推理规则,用以对数据库进行操作,是系统动态特性的描述。

数据完整性约束:描述数据结构内数据间的语法、词义联系、他们之间的制约和依存关系,以及数据动态变化的规则,保证数据的正确性、有效性和相容性。



主要的数据模型

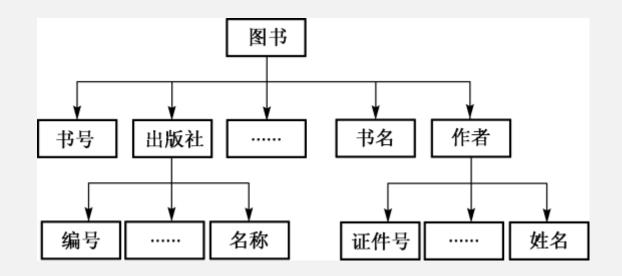
数据管理发展过程中产生过三种数据模型

- ◆ 层次模型
- ◆ 网状模型
- ◆ 关系模型

三种模型是按其数据结构而命名。



层次模型

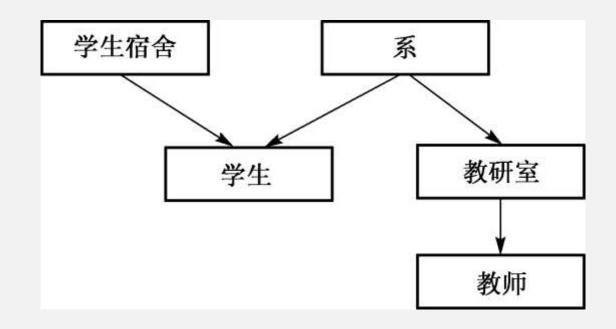


将数据组织成一对多关系的结构;

采用关键字来访问其中每一层次的每一部分。



网状模型



它用连接指令或指针来确定数据间的显式连接关系; 具有多对多类型的数据组织方式。



网状模型和层次模型很好地解决了数据的集中和共享问题

在数据独立性和抽象级别上仍有很大欠缺。

用户进行数据存取时,需要明确数据的存储结构,指出存取路径。



关系模型

用二维表的形式表示实体和实体间联系的数据模型。 表内主键区分记录,表间外键发生联系。

_		│ 关系模式↓
表 1	1	十 经 棹 元
1X		人小门大小

模式名₽	构成↩		
学生₽	(<u>学号</u> 、姓名、性别、班级、住址)↩	ŀ	
成绩₽	(<u>学号、课程号</u> 、成绩)↩		
课程₽	(<u>课程号</u> 、课程名、任课教师、学分)。	ŀ	

主 辑 ↓ 表 1.3 课程关系↓

课程号₽	课程名↩	任课教师↩	学分₽	4
4-01₽	大学英语₽	马丽↩	1₽	÷
4-02₽	大学物理↩	张远↩	1₽	4
41-03₽	大学语文₽	周明国₽	1₽	4
.1-04₽	高等数学₽	韩军₽	2₽	4
.1-05₽	计算机基础₽	王华₽	3₽	٦

表 1.2 学生关系

Е						
=	学	号→	姓名↩	性别↩	班级↩	住址↩
	20100	001119₽	张瑜婷₽	女₽	新闻一班↩	2-503₽
	20100	001120₽	马云飞↩	男₽	新闻一班↩	7-306₽
	20090	001121₽	周婷婷↩	女₽	新闻一班↩	2-503₽
		₽	٠	₽	٠٥	₽

外键

表 1.4 成绩关系

外键

_			
X	学号↩	课 程号 ₹	成绩↩
201	10001119₽	1-01↔	85₽
201	10001119₽	1-03₽	90₽
201	10001119₽	1-04₽	87₽
200	09001120₽	1-01₽	80₽
200	09001120₽	1-02₽	67₽
200	09001120₽	1-04₽	79₽
200	09001120₽	1-05₽	89₽
	٠	٠	



数据结构简单 (二维表格);

以关系运算理论、关系模式设计理论作为理论基础。



4.5 关系数据库



关 系 数 据 库

基本概念

关系: 在关系模型中,基本数据结构被限制为二维表格。因此,数据在用户观点下的逻辑结构就是一张二维表,每一张二维表称为一个关系。

编号₽	姓名↩	基本工资₽	工龄工资₽	扣除₽	实发工资₽	₽
101₽	刘安₽	1520.00₽	1532. 00₽	545.00₽	2507. 00₽	٦
102₽	陈林↩	1426.00₽	1524.00₽	530.50₽	2420. 50₽	٦
103₽	赵刚↩	1388.00₽	1525.00₽	540.50₽	2373. 50₽	٦
201₽	刘达₽	1388.00₽	1515.00₽	533.00₽	2370.00₽	ته
202₽	陆海₽	1476.00₽	1512.00₽	522.50₽	2466. 50₽	ته
203₽	李万↩	1698.00₽	1527. 00₽	560.00₽	2665. 00₽	ته



元组: 表中的行。一行为一个元组,对应存储文件中的一个记录值。

属性: 表中的列,每一列有一个属性名。属性值相当于记录中的数据项或者字段值。

候选关键字: 能够唯一地标识一个元组的属性或属性组合。

主关键字: 选择作为元组区分标志的侯选关键字。

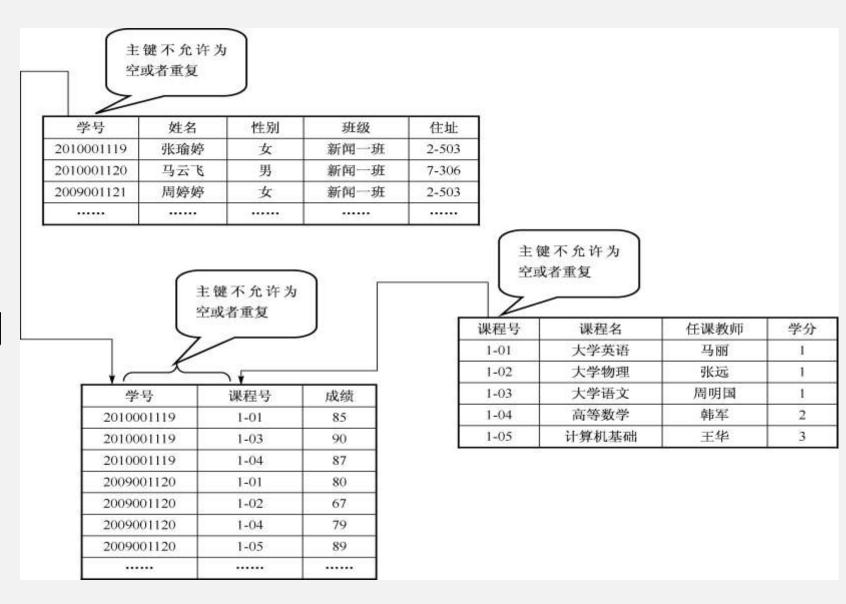
外关键字: 一个关系中的属性或属性组合并不是该关系的关键字, 但它们是另外一个关系的关键字。



完整性约束

三类规则

- (1) 实体完整性规则
- (2) 参照完整性规则
- (3) 用户定义完整性规则

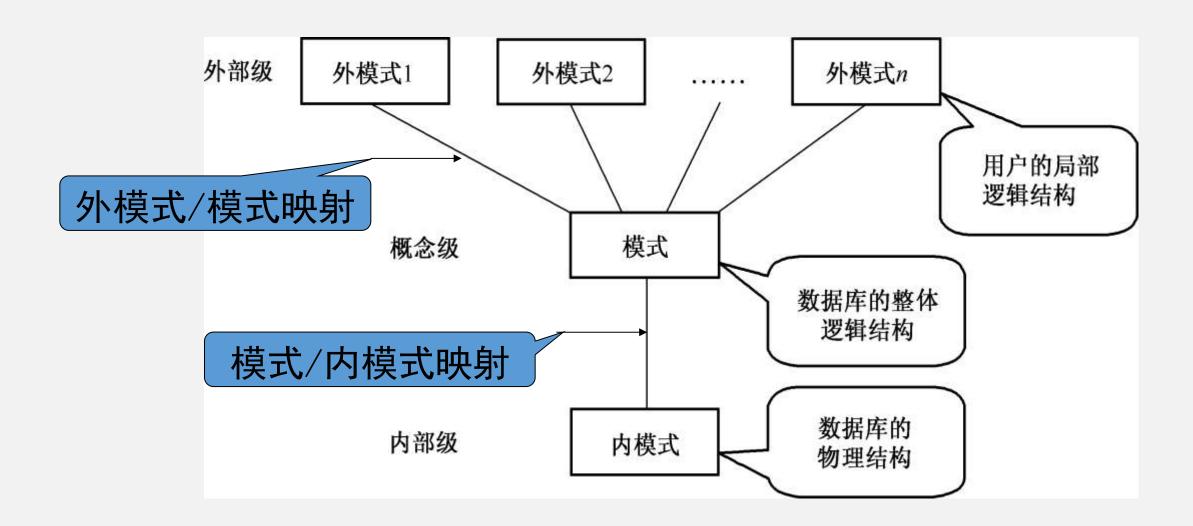




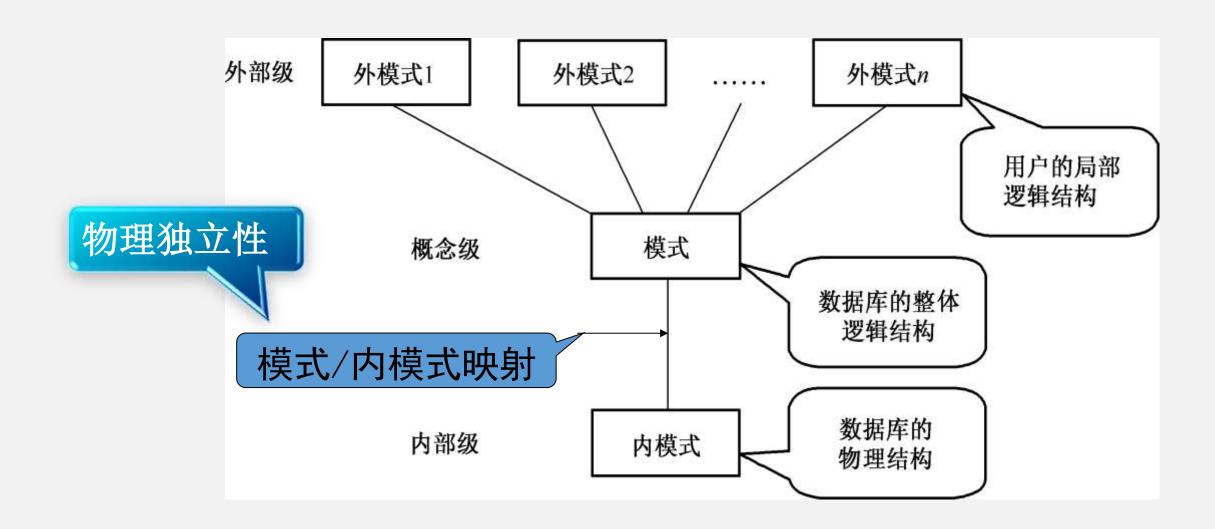
关系数据体系结构

关系数据库系统为了保证数据的逻辑独立性和物理独立性, 在体系上采用三级模式和两种映射结构。

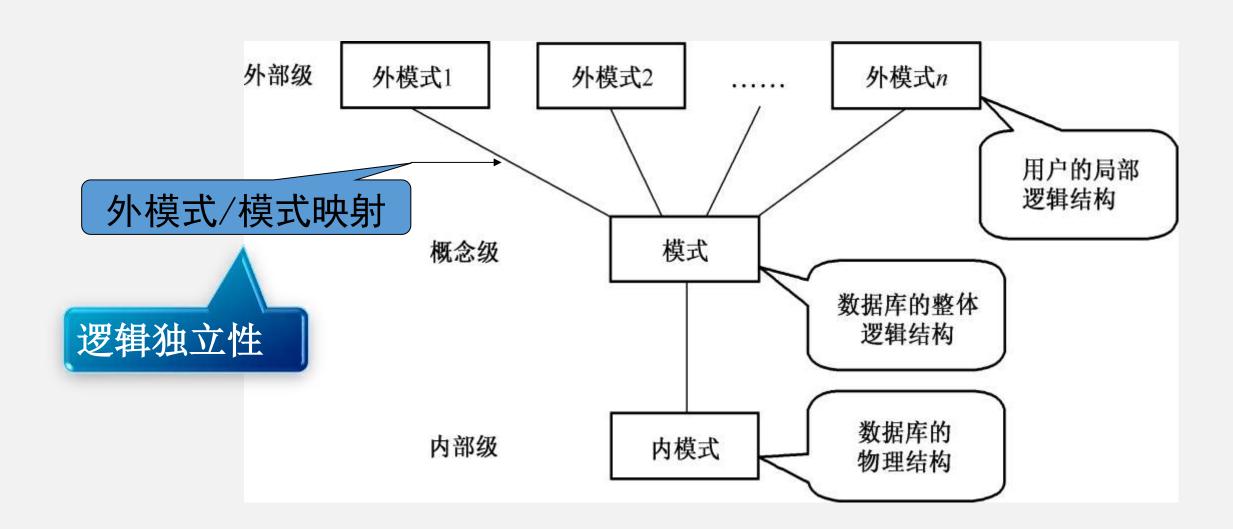














4.6 基本的关系运算



基本的关系运算

传统的集合运算

并、差、交、笛卡儿积运算和除。

并运算

并:设关系R和S具有相同的关系模式,R和S的并是由属于R或属于S的元组构成的集合,记为RUS。

 $R \cup S = \{t \mid t \in R \lor t \in S\}$ t是元组变量,R和S的元数相同。

作用:关系的合并;

要求:相同关系模式



例 设关系R、S如下:

R	(A,	В,	C)
	a	1	C
	d	4	f
	C	3	d
S	(A,	В,	C)
	Ъ	2	a
	d	4	f

(A,	В,	C)
a	1	c
b	2	a
c	3	d
d	4	f

 $R \cup S$:



差运算

差:设关系R和S具有相同的关系模式,R和S的差是由属于R但不属于S的元组构成的集合,记为R-S。

 $R-S=\{t \mid t \in R \land t \notin S\}$ t是元组变量,R和S的元数相同

作用:去掉重复的记录;

要求:相同关系模式



例 设关系R、S如下:

R	(A,	В,	C)		(A,	Β,	C)
	a	1	С	R-S:		1	
	d	4	f			1 3	
	c	3	d				
S	(A,	В,	C)		(A,	В.	C)
	Ъ	2	a	S-R:		<i>D</i> ,	
	d	4	f		b	2	a



交运算

交:关系R和S的交是由既属于R又属于S的元组构成的集合,记为R∩S,这里要求R和S定义在相同的关系模式上。形式定义如下:

 $R \cap S = \{t \mid t \in R \land t \in S\}$ t是元组变量,R和S的元数相同。

作用: 获取重复的记录;

要求:相同关系模式



例 设关系R、S如下:

R	(A,	В,	C)
	a	1	С
	d	4	f
	c	3	d
S	(A,	В,	C)
	Ъ	2	a
	d	4	f

$$R \cap S: \frac{(A, B, C)}{d \quad 4 \quad f}$$



笛卡尔积运算

笛卡儿积:设关系R和关系S的元数分别为r和s。定义R和S的笛卡儿积R×S是一个(r+s)元的元组集合,每个元组的前r个分量 (属性值)来自R的一个元组,后s个分量是S的一个元组,记为 R×S。形式定义如下:

$$R \times S = \{t \mid t = < t_r, t_s > \land t_r \in R \land t_s \in S\}$$

作用: 全组合连接, 形成新关系;



例	R	(A,	В,	<u>C)</u>
		a	1	С
		b	2	f

R×S:

(R. A,	В,	С,	S.A,	BB)
a	1	c	aa	11
a	1	c	bb	22
b	2	\mathbf{f}	aa	11
b	2	f	bb	22



专门的关系运算

包括:选择、投影和连接运算。

选择

从一个关系R中选出满足条件表达式F的元组,构成一个新关系,

作用:从一个表中选出满足条件的行,模式不变。



编号 101 102 103	姓名 武海云 张 成 马运琪	基本工资 2520.00 2426.00 2388.00	工龄工资 2532,00 2524,00 2525,00	扣除 545.00 530.50 540.50	实发工资 4507.00 4420.50 4373.50
201	周洲	2388.00	2515.00	533.00	4370.00
202	马玉涛	2476.00	2512.00	522.50	4466. 50
203	张锦强	2698.00	2527.00	560.00	4665.00
301	刘山雨	2900.00	2400.00	550.00	4750.00
301	金玉峰	2850.00	2700.00	500.00	5050.00

$\sigma_{$ 实发工资 > 4700 $}(gzb)$ select * from gzb where 实发工资 > 4700

编号	姓名	基本工资	工龄工资	扣除	实发工资
301	刘山雨	2900.00	2400.00	550.00	4750.00
301	金玉峰	2850.00	2700.00	500.00	5050.00



投影

从一个关系R中选出属性(组)A,构成一个新关系,记为口A(R)。 投影 属性列表 关系

作用:选择若干列,构成一个新表。模式改变。



编号	姓名	基本工资	工龄工资	扣除	实发工资
101	武海云	2520.00	2532.00	545.00	4507.00
102	张 成	2426.00	2524.00	530. 50	4420.50
103	马运琪	2388.00	2525.00	540.50	4373.50
201	周洲	2388.00	2515.00	533.00	4370.00
202	马玉涛	2476.00	2512.00	522.50	4466. 50
203	张锦强	2698.00	2527.00	560.00	4665.00
301	刘山雨	2900.00	2400.00	550.00	4750.00
301	金玉峰	2850.00	2700.00	500.00	5050.00

∏编号,姓名,实发工资(gzb)

select 编号,姓名,实发工资 from gzb

编号	姓名	实发工资
101	武海云	4507.00
102	张 成	4420.50
103	马运琪	4373.50
201	周 洲	4370.00
202	马玉涛	4466.50
203	张锦强	4665.00
301	刘山雨	4750.00
301	金玉峰	5050.00



连接

将多张模式不同的表通过模式合并构造一张新表。

θ连接: 先对两个关系做笛卡儿积生成一个新的关

系, 然后在新的关系上做选择操作。

自然连接:除去重复属性的等值连接。



θ-连接,记为: R ⋈S,对R和S先做笛卡尔积运算,然后选择满足θ条件的行。 θ

例	R1 (A1,	A2)	R2 (B1,	A2)
	a1	1	b1	1
	a2	2	b2	1
	a3	1	b3	2
	a4	2	b4	3



自然连接

首先对R和S进行笛卡尔积,然后选择公共字段相等的行,最后去除重复属性。

(A, B, C)	
R (A, B) S (B, C) a1 1 b1	
$\overline{a1}$ $\overline{1}$ $\overline{b1}$ $a1$ 1 $b2$	
$a2 1 1 b2 R \bowtie S = a2 1 b1$	/
$\mathbf{a} 2 2 \mathbf{b} 2 \mathbf{a} 2 1 \mathbf{b} 2$	
a3 1 a2 b2	
a3 1 b1	
a3 1 b2	



关系 S

SNO	SNAME	AGE	SEX	SDEPT
2014112001	卢雨轩	18	女	计算机
2014112009	江南	19	男	计算机
2014108093	韩晓云	18	女	新闻
2014102085	刘流	18	男	外语
2014102090	郑重	19	男	外语

关系 C

CNO	CNAME	CDEPT	TNAME
101	组成原理	计算机	张强
103	计算机网络	计算机	周明
201	新闻史	新闻	李莉
202	新闻写作	新闻	范梅
305	英语阅读	外语	杨丽华

关系 SC

SNO	CNO	SCORE
2014112001	101	78
2014112001	103	93
2014112009	101	82
2014112009	103	85
2014108093	201	78
2014108093	202	90
2014102085	305	83

求S⋈SC⋈C



SMSC

关系 S

2 47 4 7				
SN0	SNAME	AGE	SEX	SDEPT
2014112001	卢雨轩	18	女	计算机
2014112009	江南	19	男	计算机
2014108093	韩晓云	18	女	新闻
2014102085	刘流	18	男	外语
2014102090	郑重	19	男	外语

关系 SC

/C/N 00		
SN0	CNO	SCORE
2014112001	101	78
2014112001	103	93
2014112009	101	82
2014112009	103	85
2014108093	201	78
2014108093	202	90
2014102085	305	83

- ❖ 做笛卡尔积,得到一张8 列,35行的新表;
- ❖ 选择公共字段相等的行, 得到一张8列,7行的新表;

SNO	CNO	SCORE	SNO	SNAME	AGE	SEX	SDEPT
2014112001	101	78	2014112001	卢雨轩	18	女	计算机
2014112001	103	93	2014112001	卢雨轩	18	女	计算机
2014112009	101	82	2014112009	江南	19	男	计算机
2014112009	103	85	2014112009	江南	19	男	计算机
2014108093	201	78	2014108093	韩晓云	18	女	新闻
2014108093	202	90	2014108093	韩晓云	18	女	新闻
2014102085	305	83	2014102085	刘流	18	男	外语



❖ 去掉重复字段,得到一张7列,7行的新表;

SN0	CNO	SCORE	SNAME	AGE	SEX	SDEPT
2014112001	101	78	卢雨轩	18	女	计算机
2014112001	103	93	卢雨轩	18	女	计算机
2014112009	101	82	江南	19	男	计算机
2014112009	103	85	江南	19	男	计算机
2014108093	201	78	韩晓云	18	女	新闻
2014108093	202	90	韩晓云	18	女	新闻
2014102085	305	83	刘流	18	男	外语



❖ 7列7行新表和C做自然连接运算

SNO	CNO	SCORE	SNAME	AGE	SEX	SDEPT
2014112001	101	78	卢雨轩	18	女	计算机
2014112001	103	93	卢雨轩	18	女	计算机
2014112009	101	82	江南	19	男	计算机
2014112009	103	85	江南	19	男	计算机
2014108093	201	78	韩晓云	18	女	新闻
2014108093	202	90	韩晓云	18	女	新闻
2014102085	305	83	刘流	18	男	外语

关系 C

CNO	CNAME	CDEPT	TNAME
101	组成原理	计算机	张强
103	计算机网络	计算机	周明
201	新闻史	新闻	李莉
202	新闻写作	新闻	范梅
305	英语阅读	外语	杨丽华



❖ 最终结果:一张新表,10列10行

SN0	CNO	SCORE	SNAME	AGE	SEX	SDEPT	CNAME	CDEPT	TNAME
2014112001	101	78	卢雨轩	18	女	计算机	组成原理	计算机	张强
2014112001	103	93	卢雨轩	18	女	计算机	计算机网络	计算机	周明
2014112009	101	82	江南	19	男	计算机	组成原理	计算机	张强
2014112009	103	85	江南	19	男	计算机	计算机网络	计算机	周明
2014108093	201	78	韩晓云	18	女	新闻	新闻史	新闻	李莉
2014108093	202	90	韩晓云	18	女	新闻	新闻写作	新闻	范梅
2014102085	305	83	刘流	18	男	外语	英语阅读	外语	杨丽华



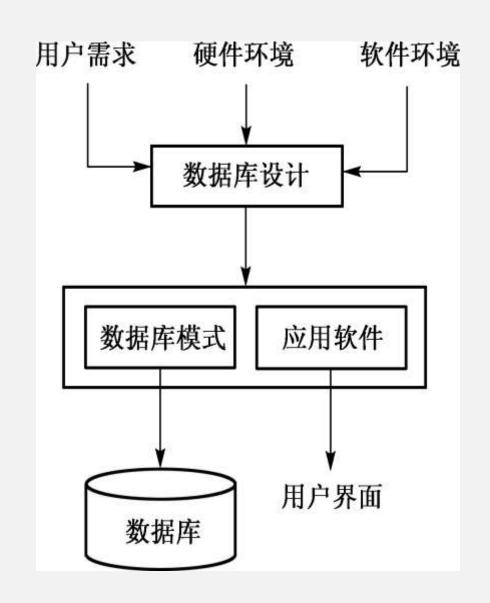
4.7 数据库设计



数据库设计

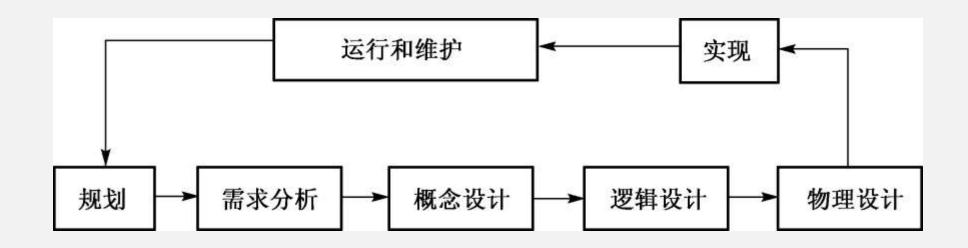
数据库设计的含义

数据库设计以DBMS为基础, 进行数据库的结构特性设计 和数据库的行为特性设计。 其中数据库的结构特性设计 起着关键作用。





数据库设计的过程



根据生命周期思想数据库设计过程可分为:规划、需求分析、概念设计、逻辑设计、物理设计、实现、运行和维护7个阶段。



规划

目的:确定系统名称、范围,确定系统的目标功能和性能,确定系统所需的资源,估算预期效益和开发成本,确定开发计划和开发进度。

两个阶段:系统调查和可行性分析。

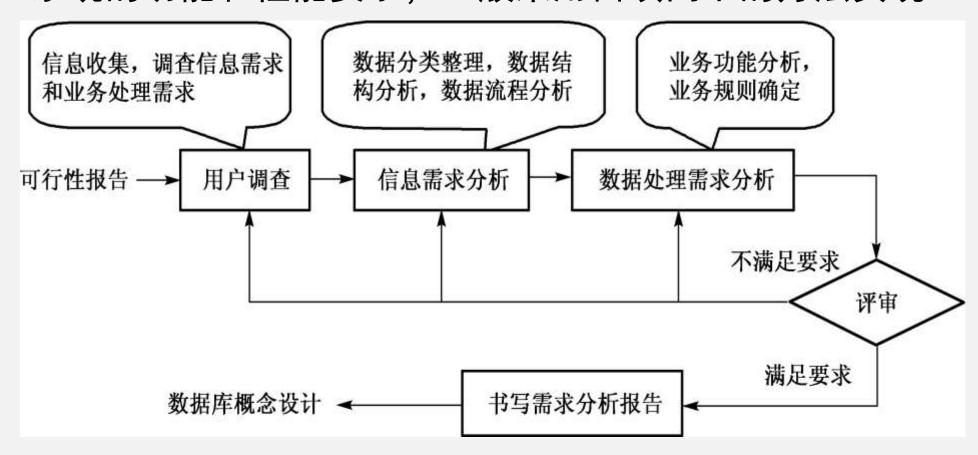
可行性分析主要考虑四个要素:

- > 经济
- > 技术
- ▶ 社会环境
- > 人

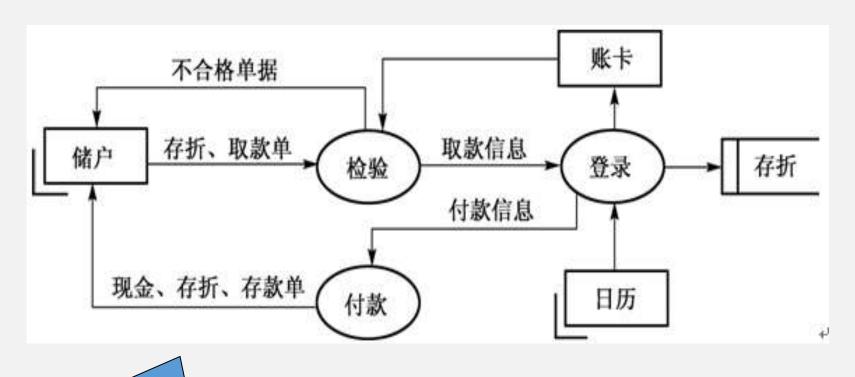


需求分析

以用户调查为基础,通过分析和综合,逐步明确用户对新 系统的功能和性能要求,一般采用自顶向下的方法实现。

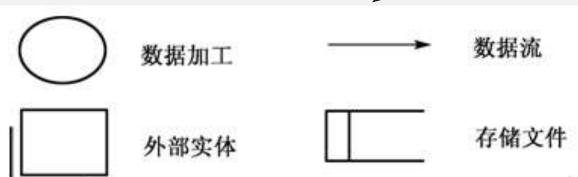






数据流图的构成元素

需求分析一般采用 数据流图描述数据 处理过程

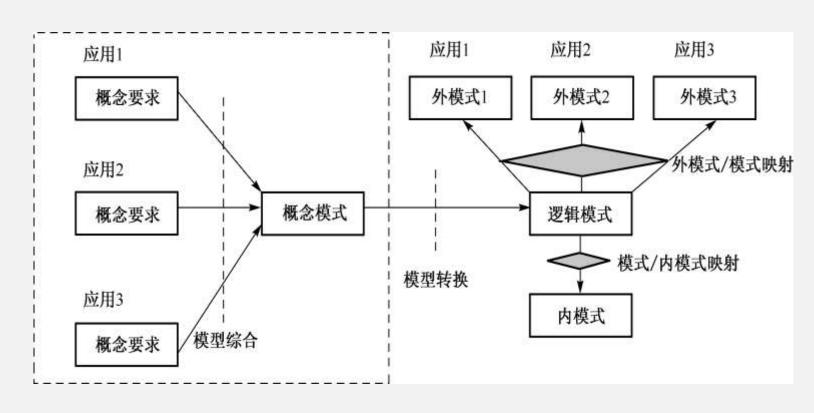




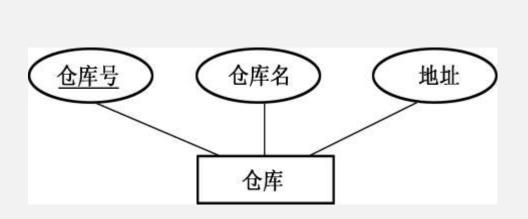
概念结构设计

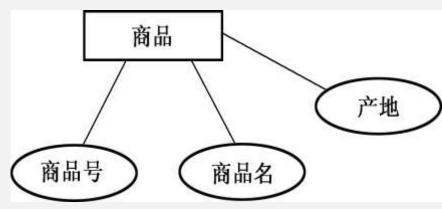
根据需求分析阶段产生的需求分析报告,产生能够反映信息需求的数据库概念结构,即数据库的逻辑模型。概念模型不依赖于计算机

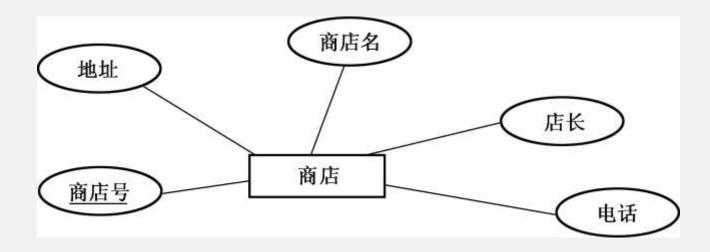
系统和具体的DBMS。



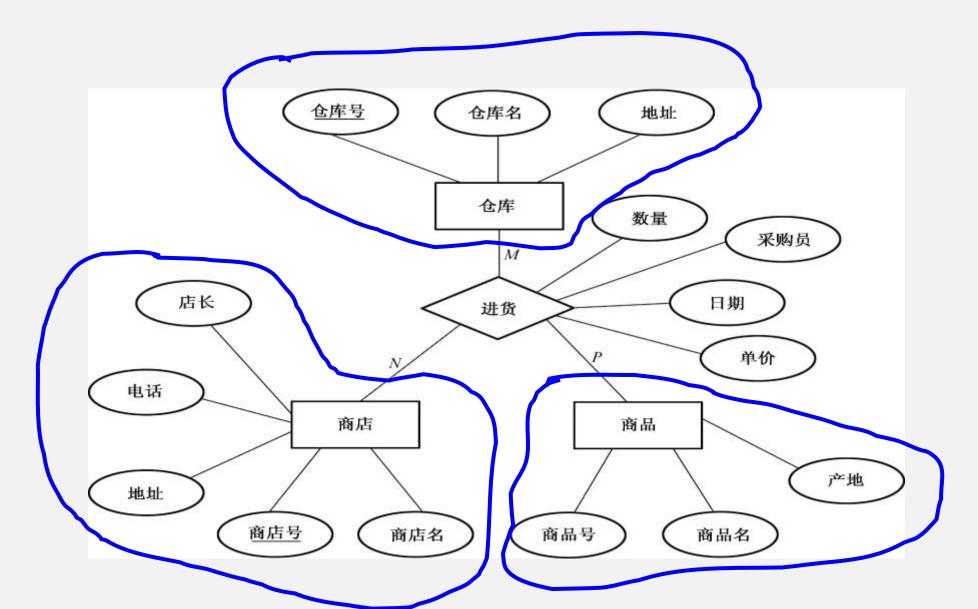








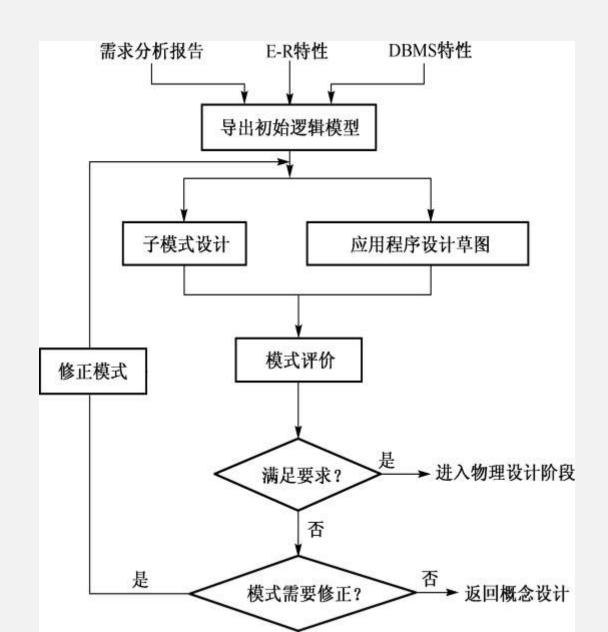






逻辑结构设计

将概念数据模型设计成数 据库能够支持的某种逻辑 模型。为各种数据处理应 用领域产生相应的逻辑子 模式. 并使其在功能、性 能、完整性约束、一致性 和可扩充性等方面均满足 用户的需求。





模型转换:把E-R图转换成关系模式的集合。

两部分转换:实体型的转换和联系的转换。

实体类型的转换:将每个实体类型转换成一个关系模式,实体的属性即为关系模式的属性,实体标识符为关系模式的键。

联系的转换

二元联系的转换

一元联系的转换

三元联系的转换



联系是1:1:在两个实体类型转换成的任意一个关系模式的属性中加入另一个关系模式的键 和联系类型的属性。

二元联系

联系是1:N: 在N端实体类型转换成的关系模式中加入1端实体类型的键和联系类型的属性。

联系是M:N:联系转换成关系模式,属性为两端实体类型的键加上联系类型的属性,而键为两端实体键的组合。



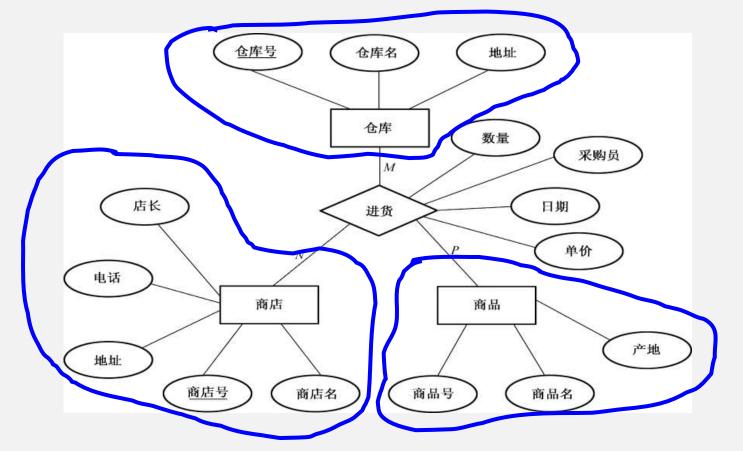
一元联系

同二元联系

三元联系

将三元联系转换成关系模式,其属性为三端实体类型的键加上联系类型的属性,而键为三端实体键的组合。





仓库(仓库号,仓库名,地址)

商店(商店号,商店名,店长,电话,地址)

商品(商品号,商品名,产地)

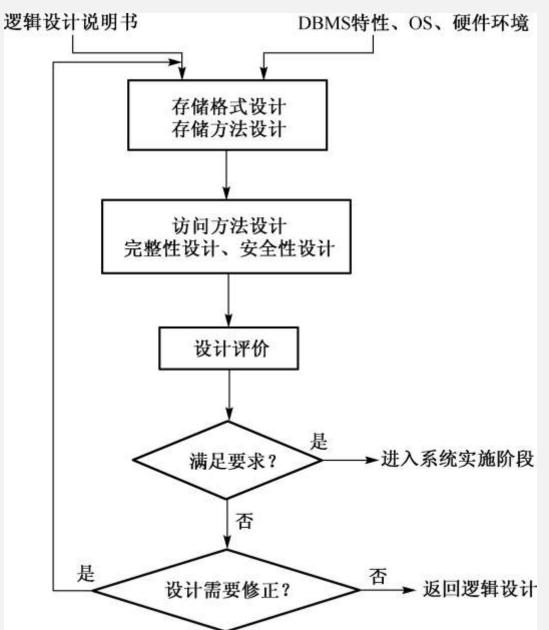
进货(<u>商店号,商品名,仓库号</u>,日期,数量,采购员,单价)



物理设计

物理设计是指在特定数据库管理系统所提供的多种存储结构和 存取方法的基础上,为逻辑数据模型选取一个最适合应用环境 的物理结构(包括文件类型、索引结构和数据的存放次序)、 存取方法和存取路径等。





存储记录的格式设计

存储方法设计

访问方法设计

完整性和安全性设计

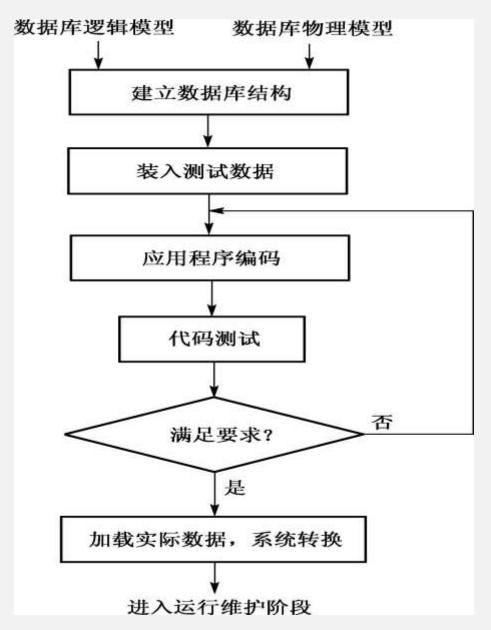
应用设计: 功能设计



系统实现

根据逻辑设计和物理设计的结果,在计算机上建立数据库,编制与调试应用程序,组织数据入库,并进行系统测试和试运行。





用DDL定义数据库结构

组织数据入库

编制与调试应用程序



运行和维护

数据库应用系统正式投入运行后,在数据库系统运行过程中必须不断地对其进行评价、调整与修改。

主要工作:

- ▶数据库的转储和恢复;
- ▶数据库安全性和完整性控制;
- ▶数据库性能的监督、分析和改进;
- ▶数据库的重组织和重构造;
- ▶系统的功能扩充与改进。