



互联网软件开发技术与实践

数据库设计

《互联网软件开发技术与实践》

课程建设小组

北京大学

二零一九年 北京



北京大学



目标

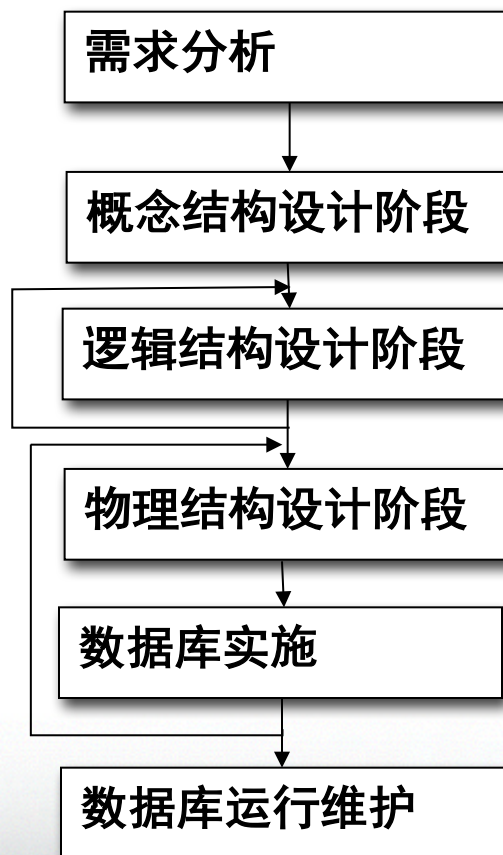
- 根据用户需求，设计合理的数据库
- 避免数据冗余、更新异常、插入异常、删除异常
- 了解E-R图、设计范式





- 数据库的设计过程可分为六个阶段：

1. 需求分析
2. 概念结构设计阶段
3. 逻辑结构设计阶段
4. 物理结构设计阶段
5. 数据库实施
6. 数据库运行与维护



北京大学



需求分析

- 需求分析是数据库设计的第一步。
- 这一阶段主要是对数据库应用系统所要处理的对象进行全面了解，了解用户需求，收集信息，提出解决问题的思路。

例如：建立“学生信息管理系统”，目的是对学院、系、班级、教研室等部门的信息进行管理，还要管理教师和学生的基本信息，以及学生选课的情况。

信息要求：系统涉及到学院、系、教研室、教师、学生、班级、课程、选课等各方面的信息。

处理要求：增加、删除、修改各数据对象中的数据，各类查询、统计及报表打印等等。



北京大學



概念结构设计

- 概念设计的目标是产生反映需求信息的数据库概念结构。
- 从现实世界到概念模型的转换是由数据库设计人员完成的
- 概念设计用**实体-联系模型(简称E-R图)**表示。
- 在E-R模型中，用于描述数据的概念主要有**实体、属性、实体型、实体集、关键字及实体之间的联系**等。





(1) 实体：客观存在并相互区别的事物。

实体可以是可触及的对象，如一个学生，一本书，一辆汽车；也可以是抽象的事件，如一堂课，一次比赛等。

(2) 属性：实体所具有的特征和性质。

如：学生实体有学号、姓名、年龄、性别等。



北京大学



(3) 码：在实体的属性当中能够唯一的标识一个实体的属性称为这个实体的关键字，也称码或键。

如：学生的学号。而学生的姓名可能有重名，不能作为学生实体的关键字(码)。

(4) 实体集：具有相同属性的同一类事物，是一个实体的集合。

如：所有的学生信息记录、所有的课程信息记录等。



北京大学



(5) 实体型：用实体名和属性名称集来描述同类实体。

如：学生(学号，姓名，年龄，性别)就是一个实体型。所有学生都可用这一实体型进行描述。

(6) 联系与联系集两个或两个以上的实体集间的关联关系的描述。

如：系与系主任、班级与班长(1:1关系)；系与教研室、班级与学生(1:n关系)；学生与课程、教师与课程(m:n关系)。

实体集之间的联系分一对一(1:1)、一对多(1:n)、多对多(m:n)3种。



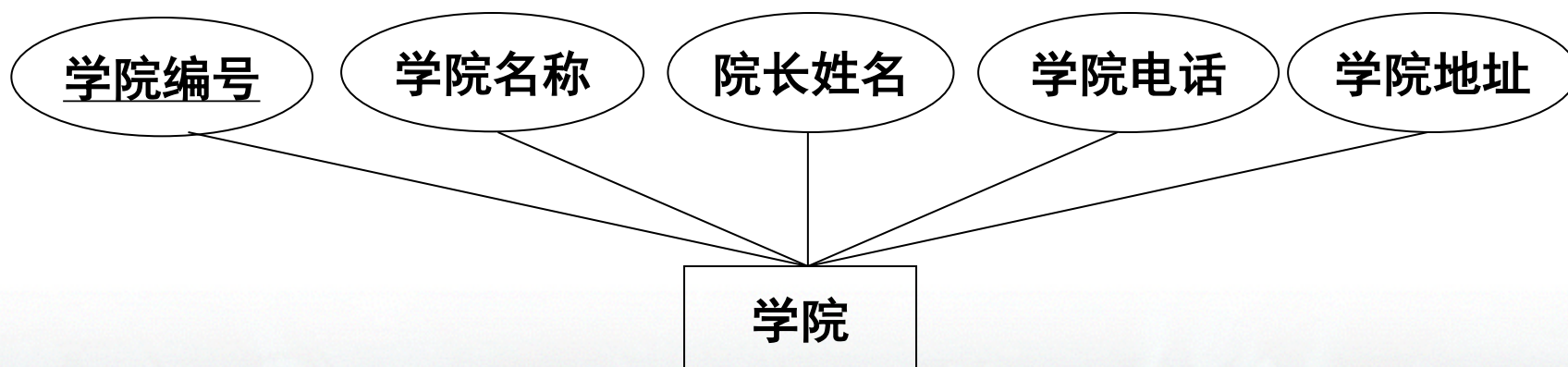
清华大学

E-R图

- E-R模型(概念模型的描述方法)，是用实体-联系图(简称E-R图)来描述的，是通过图形描述实体集、实体属性和实体集之间联系的图形。

实体集—矩形框 属性—椭圆框 联系—菱形框

例如：



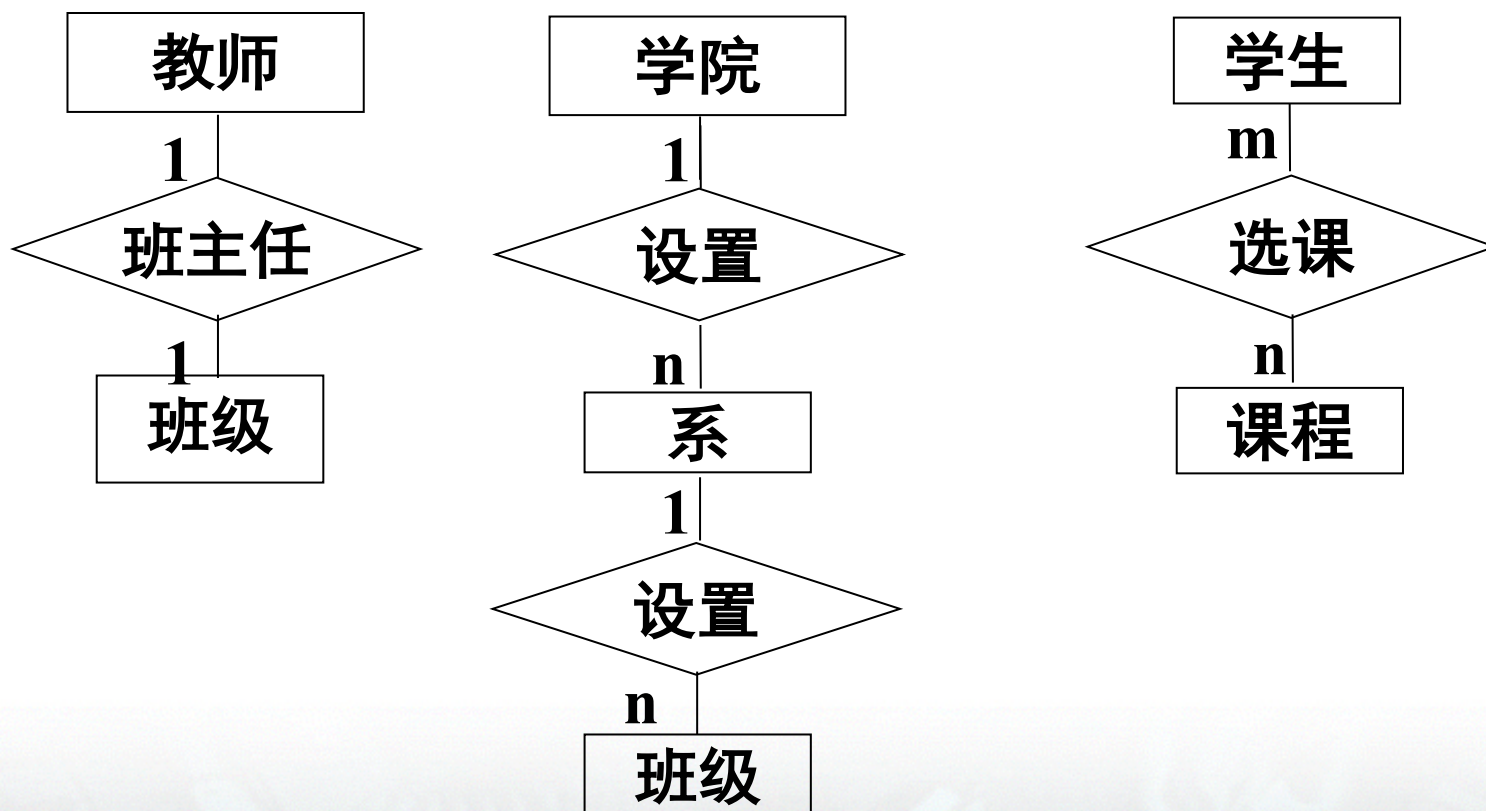
学院实体集的E-R图



北京大学



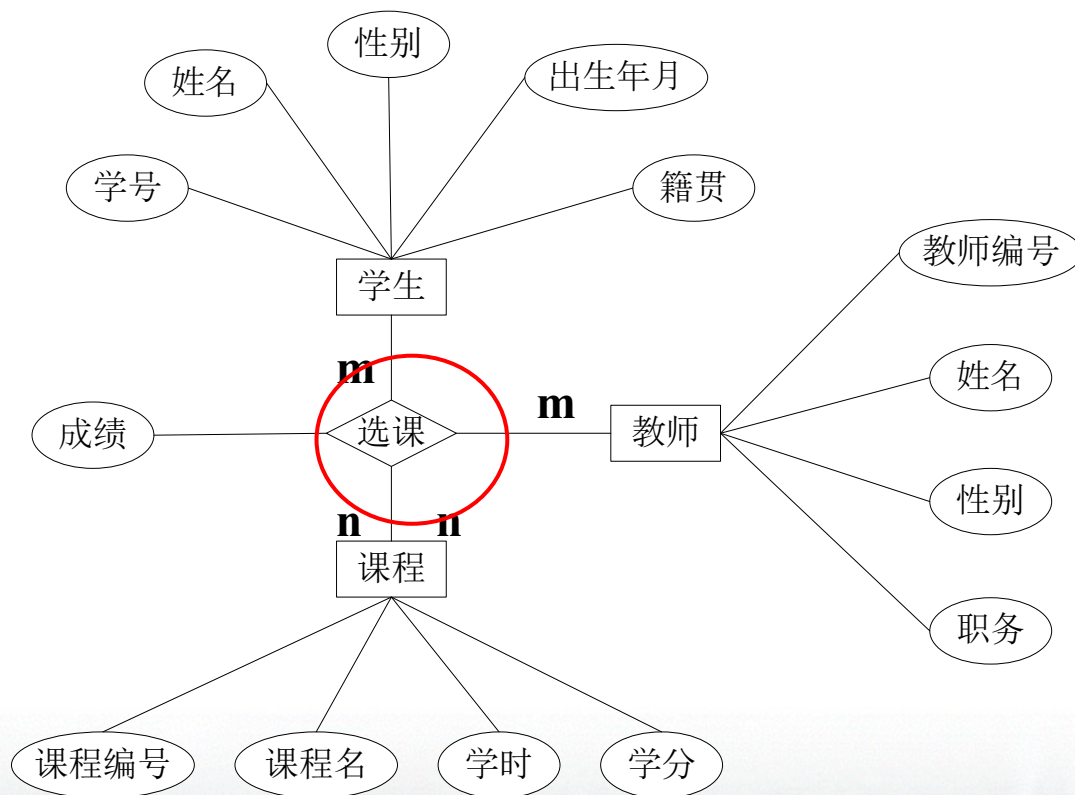
例如：教师与班级、学院与系和班级、学生和课程实体集之间联系的E-R图。



北京大学



例如：学生、教师和课程三个实体共同参与了一个选课联系，一般称这种由三类实体参与的联系为三元联系。因为某教师选定一门课讲授，学生选中同一门课要学习时，三者才会联系起来。



教师、学生与课程的E-R图



北京大学



多对多关系实际上是某两个表与第三个表的两个一对多关系，第三个表的主键包含两个字段，分别是前两个表的主键。

教师-选课 (1:n)

字段名	字段类型	字段长度	小数点	索引类型
教师编号	字符型	7		主索引
姓名	字符型	6		
性别	字符型	2		
职务	字符型	8		
教研室编号	字符型	6		外键

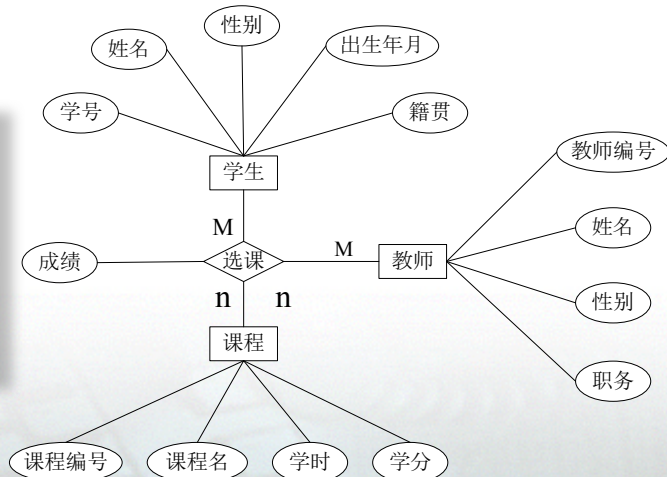
字段名	字段类型	字段长度	小数点	索引类型
学号	字符型	6		主索引
姓名	字符型	6		
性别	字符型	2		
出生年月	日期/时间型	短日期		
籍贯	字符型	50		
班级编号	字符型	8		外键

选课

字段名	字段类型	字段长度	小数点	索引类型
学号	字符型	6		普通索引
课程编号	字符型	5		普通索引
教师编号	字符型	7		普通索引
成绩	数值型	单精度	2	

课程-选课 (1:n)

字段名	字段类型	字段长度	小数点	索引类型
课程编号	字符型	5		主索引
课程名	字符型	12		
学时	数值型	短整型		
学分	数值型	短整型	0	



此图 为 3 个表与第 4 个表的联系。



北京大学

学生信息管理系统 (E-R概念模型)

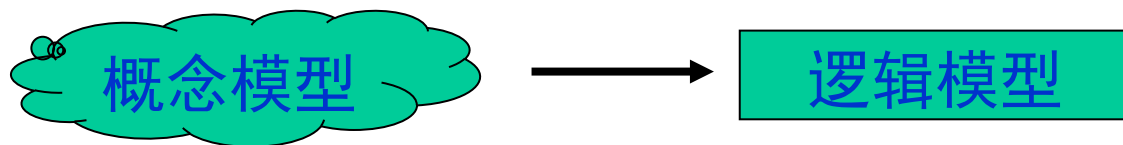


北京大学

逻辑结构设计

逻辑结构设计：将已设计好的概念模型(E-R模型)转换为与DBMS支持的数据模型相符的逻辑结构。

从概念模型到逻辑模型的转换可以由数据库设计人员完成，也可以用数据库设计工具协助设计人员完成。



北京大学



E-R图--关系模型--关系数据库的对应关系：

概念模型理论	关系模型理论	关系数据库
实体集	关系	二维表记录集
实体	元组	记录
属性	属性	字段
属性值	分量	数据项
码	主键	主键



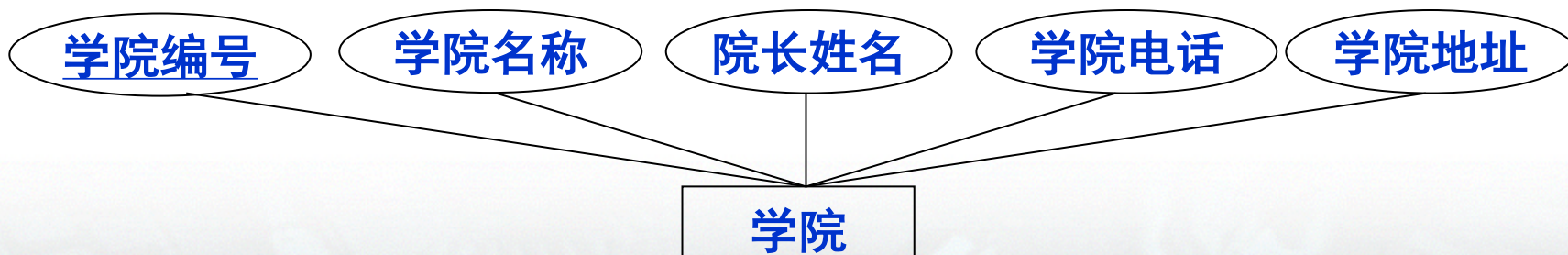


E-R模型与关系模型的转换规则：

① 一个实体型转换为一个关系模式。实体的属性就是关系的属性，实体的键就是关系的键。

例如：将学院实体转换为一个关系，其属性取原来实体的属性，利用规则1，得到关系模式：

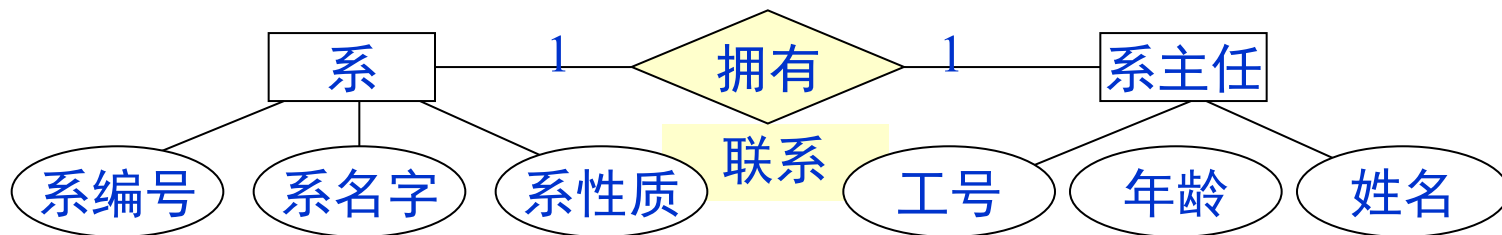
学院(学院编号,学院名称,院长姓名,学院电话,学院地址)



北京大学

② 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。

举例：两个实体：一对一联系



方式1

系(系编号,系名字,系性质)
系主任 (工号,姓名,年龄)
系与系主任(工号,系编号)

方式2

系(系编号,系名字,系性质)
系主任 (工号,姓名,年龄,系编号)

方式3

系(系编号,系名字,系性质,工号)
系主任 (工号,姓名,年龄)

方式4

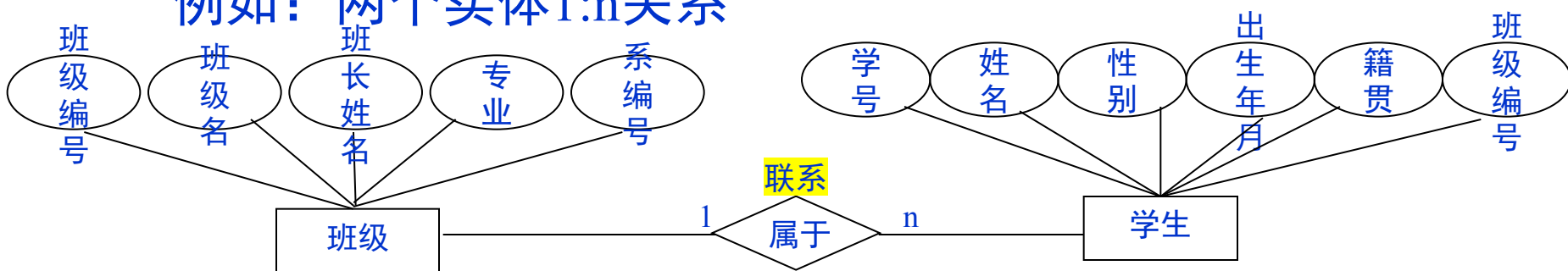
系(系编号,系名字,系性质,系主任姓名,年龄)



北京大学

③ 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。

例如：两个实体1:n关系



利用规则1将学生和班级实体各转换为一个关系，其属性取原来实体的属性，得到关系模式：

学生(学号,姓名,性别,出生年月,籍贯)

班级(班级编号, 班级名称, 班长姓名, 专业, 系编号)

由于班级与学生之间存在一对多的联系，利用规则3可将这种联系合并到n端，即将班级关系的码加入到学生关系中，得到最终结果：学生(学号,姓名,性别,出生年月,籍贯,班级编号)



北京航空航天大学



④ 实体之间多对多的联系转换为一个独立的关系模式。与该联系相关系的各实体的码及联系本身的属性均转换为此关系模式的属性，而关系的码为各实体的码的组合。

⑤ 三个或以上的实体间的一个多元联系可以转换为一个关系模式。

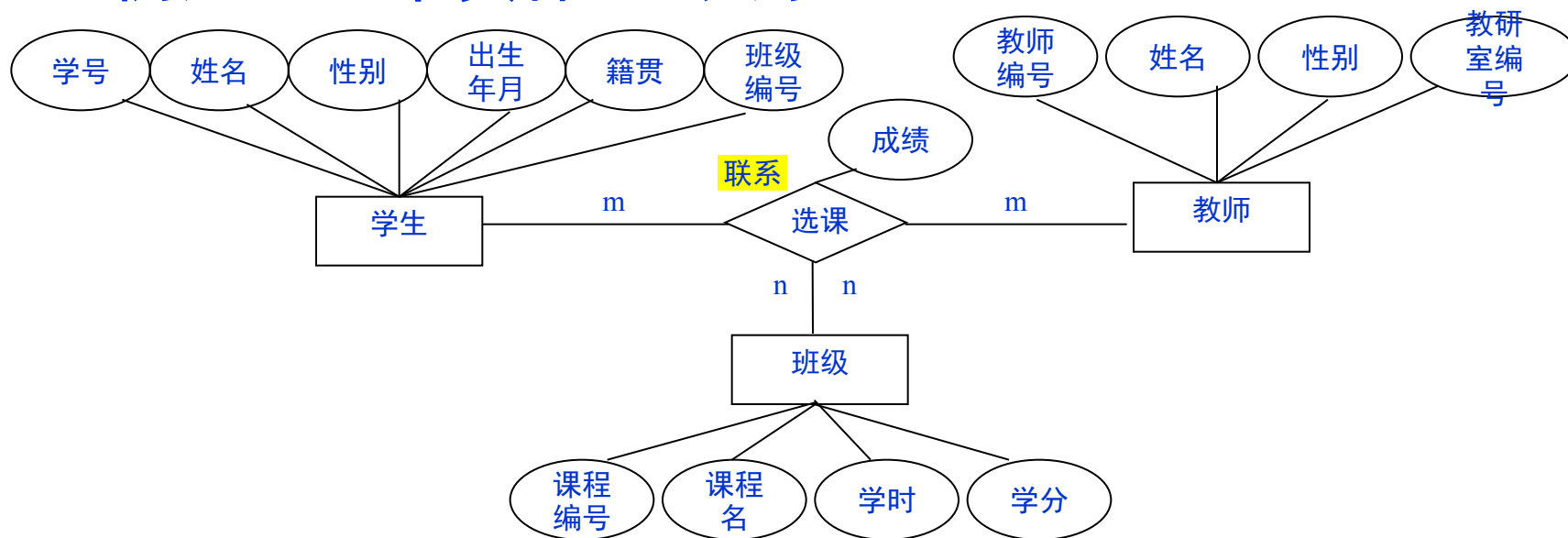
⑥ 具有相同键的关系模式可以合并。



清华大学



例如：三个实体m:n关系



使用规则1：三个实体集转换为关系；

使用规则4和5：为联系单独建立一个关系，即将学生—课程、教师—课程实体之间多对多的多元“选课”联系转换为一个独立的关系模式。



北京大学



实体之间多对多的多元“选课”联系转换为一个独立的关系模式如下：

学生(学号,姓名,性别,出生年月,籍贯,班级编号)

教师(教师编号,姓名,性别,职务,教研室编号)

课程(课程编号,课程名,学时,学分)

选课(学号,课程编号,教师编号,成绩)

(新关系包含两组多对多实体的主键及其联系自身的属性)



北京大学



- ✓学院(学院编号,学院名称,院长姓名,电话,地址)
- ✓系(系编号,系名称,系主任,学院编号)
- ✓教研室(教研室编号,教研室名称,系编号)
- ✓教师(教师编号,姓名,性别,职务,教研室编号)
- ✓班级(班级编号,班级名称,班长姓名,专业,系编号)
- ✓学生(学号,姓名,性别,出生年月,籍贯,班级编号)
- ✓课程(课程编号,课程名,学时,学分)
- ✓选课(学号,课程编号,教师编号,成绩)

学生信息管理系统(逻辑模型)



北京大学



范式及其关系

关系规范化理论简称范式，它提供了判别关系模式设计的优劣标准，为数据库设计提供了严格的理论基础。

使用范式表示关系模式满足规范化的等级，满足最低要求的为第一范式，在第一范式的基础上满足进一步要求的可升级为第二范式，其余以此类推。

范式(1NF——5NF)，主要为1NF-3NF。也就是说能满足前3个范式的关系基本上是合理的。





(1) 1NF: 若一个关系模式R的**所有属性都是不可再分的基本数据项**，则该关系模式属于1NF。

例如：教师关系表如下，判断是否为第一范式，并规范教师关系。

教师关系

教师编号	姓名	系别	联系电话	
			联系电话1	联系电话2
95010	张乐	经济	1234	12345678
95011	赵希明	经济	2345	23456789
95012	李小平	经济	3456	34567890
...

修改后的教师关系

教师编号	姓名	系别	联系电话1	联系电话2
95010	张乐	经济	1234	12345678
95011	赵希明	经济	2345	23456789
95012	李小平	经济	3456	34567890
...

不符合第一范式，因为“联系电话”不是基本数据项。

属于第一范式的关系应该满足的基本条件是每个元组的每个属性中只能包含一个数据项，不能将两个以上数据项“挤入”到一个属性中。



北京大学



(2) 2NF: 若关系模式R属于1NF, 且**每个非主属性都完全依赖于主关键字**, 则该关系模式属于2NF。

例如: 学生选课成绩关系表如下, 判断是否为第二范式, 并规范学生选课关系。

学生选课成绩关系

学生编号	姓名	课程编号	课程名称	周学时	学分	成绩
20020102	刘力	101	数据库	3	3	77
20020104	刘红	102	英语	6	6	67
		

在这个关系中**学生编号和课程编号共同组成主关键字**, 其中**成绩完全依赖于主关键字**, 而**姓名却完全依赖于学号**, **课程名称、周学时、学分又完全依赖于课程编号**, 因此, 此关系**不符合第二范式**。





上述关系可能带来以下几个问题：

① 数据冗余：假设100个学生选同一门课，重复100次相同的学分。

② 更新复杂：若调整某门课的学分，与其相关的记录都要修改。

③ 删除异常：若学生已毕业，由于学生编号不存在，选课记录也要删除。

④ 插入异常：或开一门新课，还没有学生选，由于没有学生编号，课程编号、学分均无法输入。

可以将上述关系分解为三个关系：

① 学生(学生编号, 姓名, 年龄, 入校日期)

② 选课(学生编号, 课程编号, 成绩)

③ 课程(课程编号, 课程名称, 周学时, 学分)



北京大学



(3) 3NF: 若关系模式R属于1NF, 且每个非主属性都不传递依赖于主关键字, 则关系模式属于3NF。(略)

学生选课成绩的三个关系:

- ① 学生(学生编号, 姓名, 年龄, 入校日期)
- ② 选课成绩(学生编号, 课程编号, 成绩)
- ③ 课程(课程编号, 课程名称, 周学时, 学分)

其中① 和②属于第三范式, ③如果学分是依据周学时多少来决定, 那么学分就是通过周学时传递依赖于课程号。

解决方法是, 将课程关系进一步分解为两个关系。

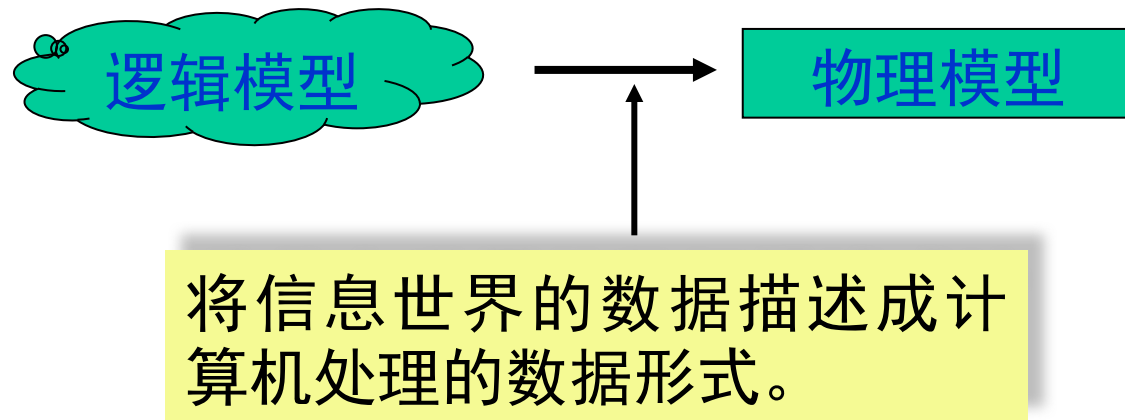
- ① 课程(课程编号, 课程名称, 周学时)
- ② 学分(周学时, 学分)



北京大学

物理结构设计

数据库物理结构设计就是为已设计好的逻辑数据模型选择最适合应用要求的物理结构。它依赖于所选择的数据库管理系统和给定的计算机硬件软件环境。





日常管理

- 用户端

- 多类用户：院领导、部门领导、系领导、普通用户等
- 用户每周及时更新自己的日程（时间、地点、事项），如（1月1日周一上午，校本部，开会）
- 对未来的日程可以修改，已过去的日程不能修改
- 各类用户可以查看所有人日程
- 可设置默认的地点和事项

- 管理端

- 各类用户类型及用户进行管理（增删改查）
- 对日程时间、地点、事项进行管理
- 设置待填写日程的周期及范围
- 记录所有的操作日志
- 统计功能



北京大學



Q&A

本讲结束 !



北京大学