

请**现场**的同学们：

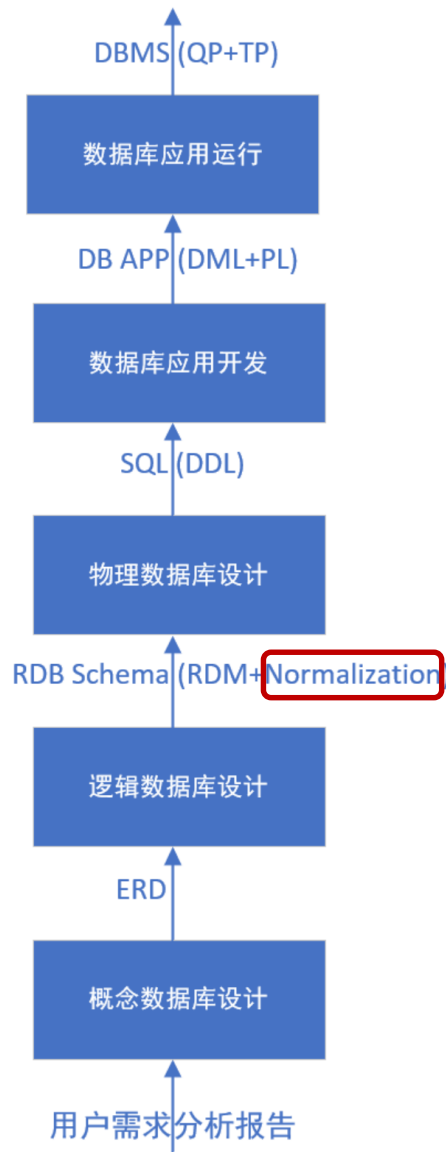
1. 打开雨课堂，点击页面右下角喇叭按钮调至静音状态

本次课程是

线上+线下 融合式教学

请**远程上课**的同学们：

1. 打开雨课堂，点击页面右下角喇叭按钮调至静音状态
2. 打开“瞩目”（会议室：182 943 865；密码：见学堂公告），进入会议室，并关闭麦克风



Outline

- DB Development Lifecycle
- Entity-relationship Modeling*
- ✈ • Database Normalization*
- Database Design

某医院病房计算机管理中心需要如下信息：

科室：科名、科地址、科电话、医生姓名

病房：病房号、床位号、所属科室名

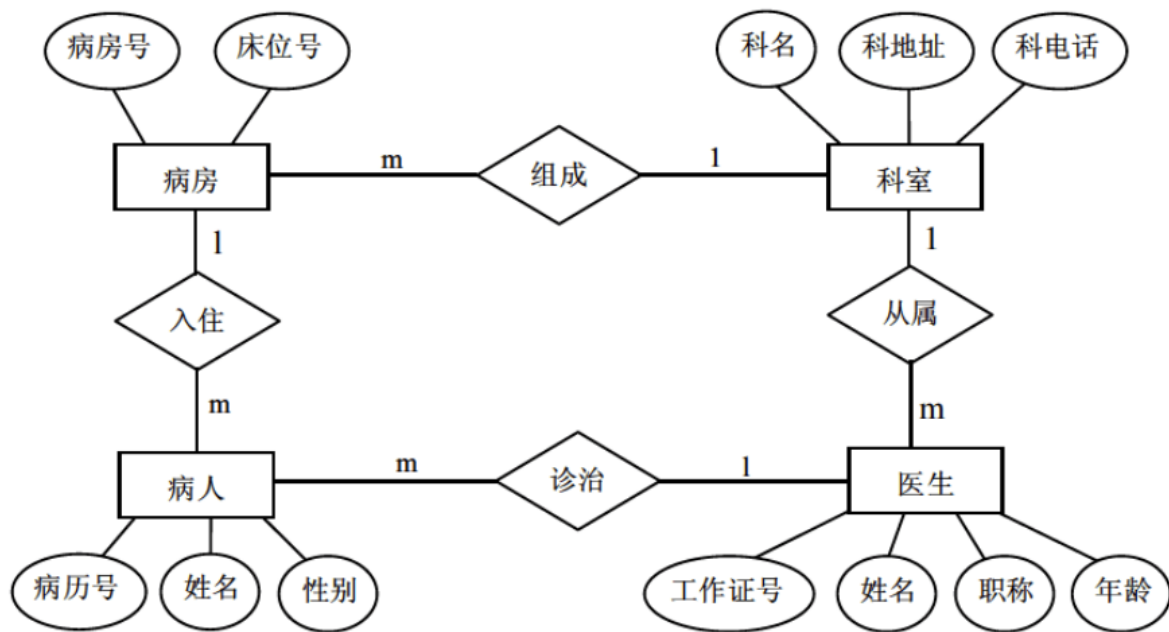
医生：姓名、职称、所属科室名、年龄、工作证号

病人：病历号、姓名、性别、主管医生、病房号

其中，一个科室有多个病房、多个医生，
一个病房只能属于一个科室，
一个医生只属于一个科室，但可负责多个病人的诊治，
一个病人的主管医生只有一个。

完成如下设计：

- (1) 设计该计算机管理系统的E-R图。
- (2) 将该E-R图转换为关系模式结构。
- (3) 指出转换结果中每个关系模式的主键。



- 对应的关系模式结构如下：

- ✓ 科室 (科名, 科地址, 科电话)
- ✓ 病房 (科室名, 病房号, 床位号)
- ✓ 医生 (工作证号, 姓名, 职称, 科室名, 年龄)
- ✓ 病人 (病历号, 姓名, 性别, 主管医生, 病房号)

一个图书借阅管理数据库要求提供下述服务：

(1) 可随时查询书库中现有书籍的品种、数量与存放位置。所有各类书籍均可由书号唯一标识。

(2) 可随时查询书籍借还情况，包括借书人单位、姓名、借书证号、借书日期和还书日期。

约定：任何人可借多种书，任何一种书可为多个人所借，借书证号具有惟一性。

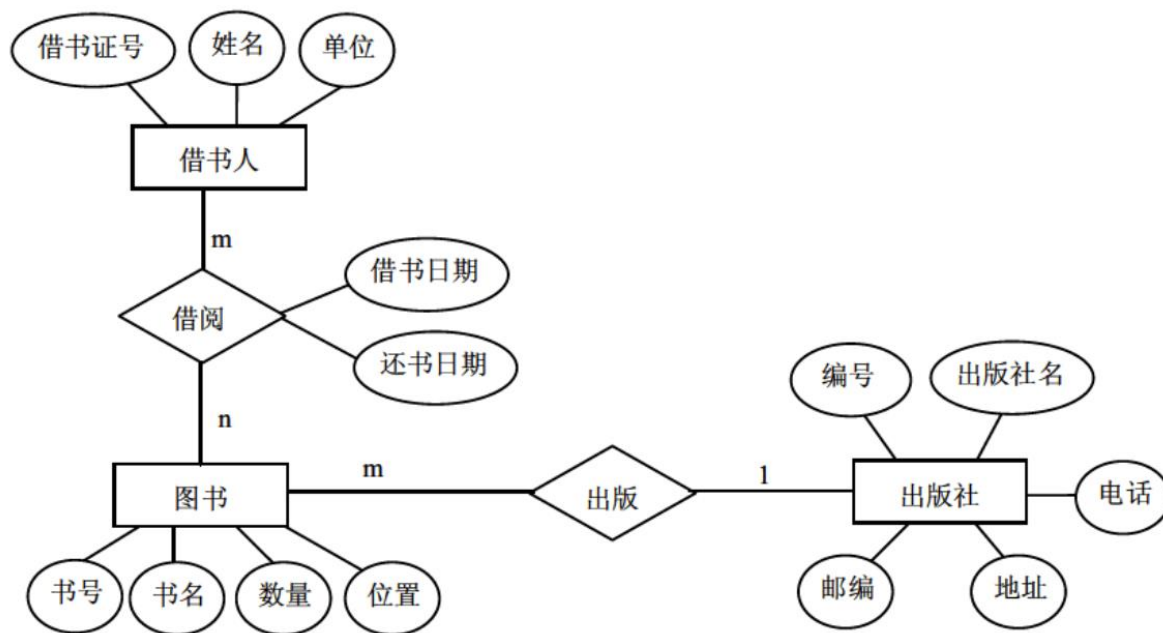
(3) 当需要时，可通过数据库中保存的出版社的电报编号、电话、邮编及地址等信息向相应出版社增购有关书籍。

约定：一个出版社可出版多种书籍，同一本书仅为一个出版社出版，出版社名具有惟一性。

根据以上情况和假设，试作如下设计：

(1) 构造满足需求的E-R图。

(2) 转换为等价的关系模式结构。



• 转换为等价的关系模式结构如下：

- ✓ 借书人（借书证号，姓名，单位）
- ✓ 图书（书号，书名，数量，位置，出版社名）
- ✓ 出版社（出版社名，电报编号，电话，邮编，地址）
- ✓ 借阅（借书证号，书号，借书日期，还书日期）

Map entities and relationships to relations

Strong entity	Create relation that includes all simple attributes.
Weak entity	Create relation that includes all simple attributes (primary key still has to be identified after the relationship with each owner entity has been mapped).
1:* binary relationship 一对多关系体现在多这边	Post primary key of entity on the "one" side to act as foreign key in relation representing entity on the "many" side. Any attributes of relationship are also posted to the "many" side.
1:1 binary relationship: (a) Mandatory participation on both sides (b) Mandatory participation on one side (c) Optional participation on both sides	Combine entities into one relation. Post primary key of entity on the "optional" side to act as foreign key in relation representing entity on the "mandatory" side. Arbitrary without further information.
Superclass/subclass relationship	See Table 17.1.
: binary relationship, complex relationship	Create a relation to represent the relationship and include any attributes of the relationship. Post a copy of the primary keys from each of the owner entities into the new relation to act as foreign keys.
Multi-valued attribute	Create a relation to represent the multi-valued attribute and post a copy of the primary key of the owner entity into the new relation to act as a foreign key.

Data Redundancy (Motivation)

增删改异常的情况

Staff Branch

staffNo	sName	position	salary	branchNo	bAddress
SL21	John White	Manager	30000	B005	22 Deer Rd, London
SG37	Ann Beech	Assistant	12000	B003	163 Main St, Glasgow
SG14	David Ford	Supervisor	18000	B003	163 Main St, Glasgow
SA9	Mary Howe	Assistant	9000	B007	16 Argyll St, Aberdeen
SG5	Susan Brand	Manager	24000	B003	163 Main St, Glasgow
SL41	Julie Lee	Assistant	9000	B005	22 Deer Rd, London

Staff

staffNo	sName	position	salary	branchNo
SL21	John White	Manager	30000	B005
SG37	Ann Beech	Assistant	12000	B003
SG14	David Ford	Supervisor	18000	B003
SA9	Mary Howe	Assistant	9000	B007
SG5	Susan Brand	Manager	24000	B003
SL41	Julie Lee	Assistant	9000	B005

Branch

branchNo	bAddress
B005	22 Deer Rd, London
B007	16 Argyll St, Aberdeen
B003	163 Main St, Glasgow

Database Tables and Normalization

- Normalization: evaluating and correcting table structures to minimize data redundancies
 - Reduces data anomalies
 - Assigns attributes to tables based on determination
- Normal forms
 - First normal form (1NF)
 - Second normal form (2NF)
 - Third normal form (3NF)

The Need for Normalization

- Used while designing a new database structure
 - Analyzes the relationship among the attributes within **each** entity
 - Determines if the structure can be improved through normalization
 - Improves the existing data structure and creates an appropriate database design
- Structural point of view of normal forms
 - Higher normal forms are better than lower normal forms

The Normalization Process

- Objective is to ensure that each table conforms to the concept of well-formed relations
 - Each table represents a single subject
 - Each row/column intersection contains only **one** value and **not** a group of values
 - No data item will be unnecessarily stored in more than one table
 - All nonprime attributes in a table are dependent on the primary key
 - Each table has no insertion, update, or deletion anomalies

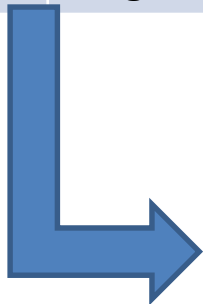
The Normalization Process (cont)

- Ensures that all tables are in at least 3NF
 - Higher forms are not likely to be encountered in business environment
- Works one relation at a time
 - Identifies the dependencies of a relation (table)
 - Progressively breaks the relation up into a new set of relations

Basic Idea

- The domain of each attribute contains only atomic values, and the value of each attribute contains only a single value from that domain

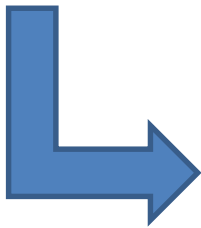
Employee	Skill	Branch	Work Location
Brown	Simulation, Database	B1	73 Industrial Way, London, UK
Harrison	Database	B1	73 Industrial Way, London, UK
Jones	Simulation, Programme	B2	114 Main Street, Dublin, Ireland



Employee	Skill	Branch	Country	City	Address Line
Brown	Simulation	B1	UK	London	73 Industrial Way
Brown	Database	B1	UK	London	73 Industrial Way
Harrison	Database	B1	UK	London	73 Industrial Way
Jones	Simulation	B2	Ireland	Dublin	114 Main Street
Jones	Programme	B2	Ireland	Dublin	114 Main Street

Basic Idea (cont)

Employee	Skill	Branch	Country	City	Address Line
Brown	Simulation	B1	UK	London	73 Industrial Way
Brown	Database	B1	UK	London	73 Industrial Way
Harrison	Database	B1	UK	London	73 Industrial Way
Jones	Simulation	B2	Ireland	Dublin	114 Main Street
Jones	Programme	B2	Ireland	Dublin	114 Main Street



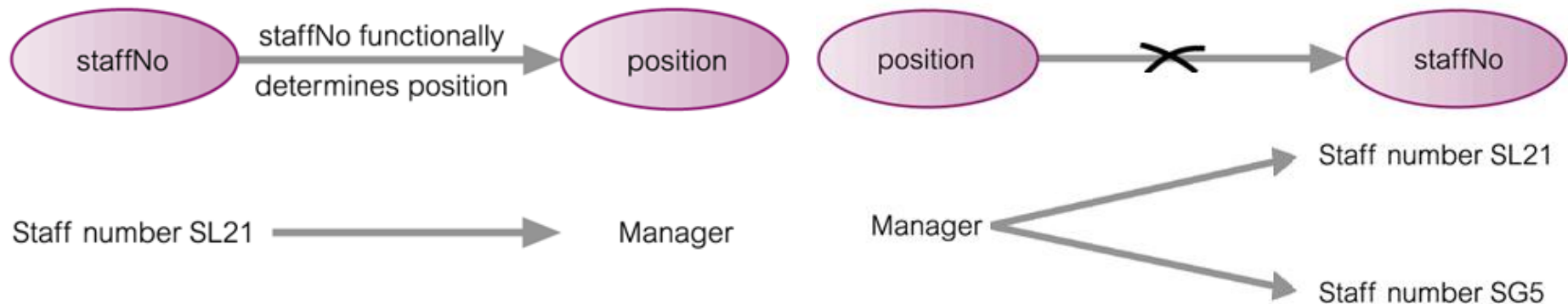
Employee	Branch
Brown	B1
Harrison	B1
Jones	B2

Branch	Country	City	Address Line
B1	UK	London	73 Industrial Way
B2	Ireland	Dublin	114 Main Street

Employee	Skill
Brown	Simulation
Brown	Database
Harrison	Database
Jones	Simulation
Jones	Programme

Functional Dependencies

- Functional dependency (函数依赖) describes relationship between attributes.
 - If A and B are attributes of relation R, B is functionally dependent on A (denoted $A \rightarrow B$), if each value of A in R is associated with exactly one value of B in R.



Full Functional Dependencies

- Full functional dependency (完全函数依赖)
 - A and B are attributes of a relation. B is fully functionally dependent on A, if B is functionally dependent on A, but not on any proper subset of A.

Staff

staffNo	sName	position	salary	branchNo
SL21	John White	Manager	30000	B005
SG37	Ann Beech	Assistant	12000	B003
SG14	David Ford	Supervisor	18000	B003
SA9	Mary Howe	Assistant	9000	B007
SG5	Susan Brand	Manager	24000	B003
SL41	Julie Lee	Assistant	9000	B005

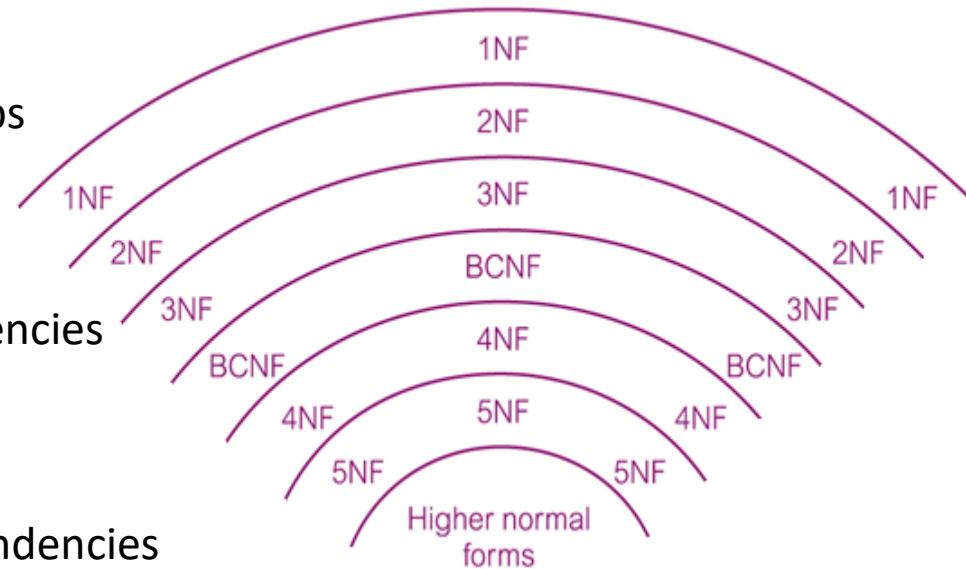
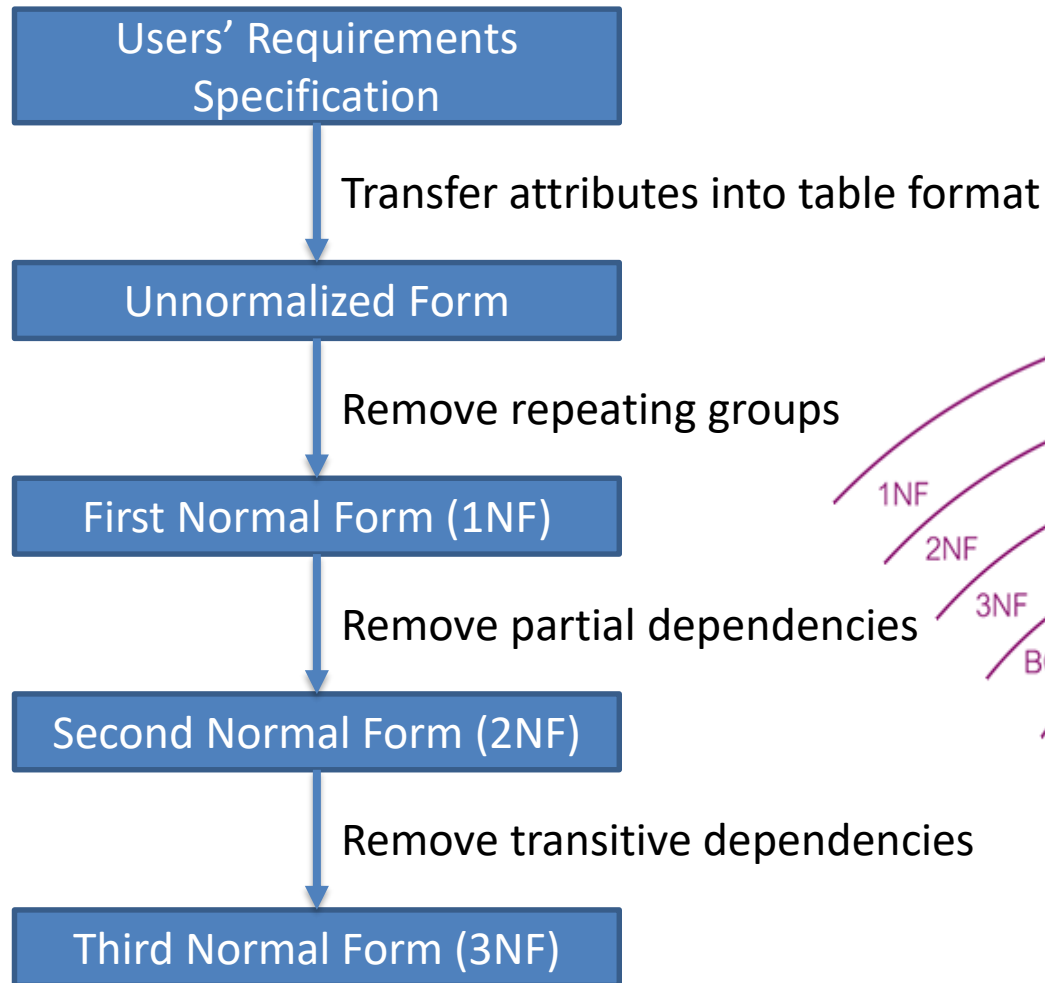
Transitive Dependencies

- Transitive dependency (依赖) describes a condition where A, B, and C are attributes of a relation such that if $A \rightarrow B$ and $B \rightarrow C$, then C is transitively dependent on A via B (provided that A is not functionally dependent on B or C).

StaffBranch

staffNo	sName	position	salary	branchNo	bAddress
SL21	John White	Manager	30000	B005	22 Deer Rd, London
SG37	Ann Beech	Assistant	12000	B003	163 Main St, Glasgow
SG14	David Ford	Supervisor	18000	B003	163 Main St, Glasgow
SA9	Mary Howe	Assistant	9000	B007	16 Argyll St, Aberdeen
SG5	Susan Brand	Manager	24000	B003	163 Main St, Glasgow
SL41	Julie Lee	Assistant	9000	B005	22 Deer Rd, London

The Process of Normalization



First Normal Form (1NF, 第一范式)

- A relation in which the intersection of each row and column contains one and only one value.
- All attributes are dependent on the primary key.

clientNo	cName	propertyNo	pAddress	rentStart	rentFinish	rent	ownerNo	oName
CR76	John Kay	PG4	6 Lawrence St, Glasgow	1-Jul-12	31-Aug-13	350	CO40	Tina Murphy
		PG16	5 Novar Dr, Glasgow	1-Sep-13	1-Sep-14	50	CO93	Tony Shaw
CR56	Aline Stewart	PG4	6 Lawrence St, Glasgow	1-Sep-11	10-June-12	350	CO40	Tina Murphy
		PG36	2 Manor Rd, Glasgow	10-Oct-12	1-Dec-13	375	CO93	Tony Shaw
		PG16	5 Novar Dr, Glasgow	1-Nov-14	10-Aug-15	450	CO93	Tony Shaw

原子性：每一个属性都不可再分
· 表中同一列的类型相同
· 每一列都不可再分

比如这里同一行里有三个元素，可以拆分成三行

clientNo	propertyNo	cName	pAddress	rentStart	rentFinish	rent	ownerNo	oName
CR76	PG4	John Kay	6 Lawrence St, Glasgow	1-Jul-12	31-Aug-13	350	CO40	Tina Murphy
CR76	PG16	John Kay	5 Novar Dr, Glasgow	1-Sep-13	1-Sep-14	450	CO93	Tony Shaw
CR56	PG4	Aline Stewart	6 Lawrence St, Glasgow	1-Sep-11	10-Jun-12	350	CO40	Tina Murphy
CR56	PG36	Aline Stewart	2 Manor Rd, Glasgow	10-Oct-12	1-Dec-13	375	CO93	Tony Shaw
CR56	PG16	Aline Stewart	5 Novar Dr, Glasgow	1-Nov-14	10-Aug-15	450	CO93	Tony Shaw

Database Con

Second Normal Form (2NF, 第二范式)

非主键对于主键完全函数依赖

- A relation that is in first normal form and every non-primary-key attribute is fully functionally dependent on the primary key.

“部分依赖于主键”
意思是：我不用主键也能确定我的键值
比如（身份证号，学号，姓名）
我用身份证号/学号都能确定姓名
则其是部分函数依赖
不符合2NF

ClientRental

clientNo	propertyNo	cName	pAddress	rentStart	rentFinish	rent	ownerNo	oName
CR76	PG4	John Kay	6 Lawrence St, Glasgow	1-Jul-12	31-Aug-13	350	CO40	Tina Murphy
CR76	PG16	John Kay	5 Novar Dr, Glasgow	1-Sep-13	1-Sep-14	450	CO93	Tony Shaw
CR56	PG4	Aline Stewart	6 Lawrence St, Glasgow	1-Sep-11	10-Jun-12	350	CO40	Tina Murphy
CR56	PG36	Aline Stewart	2 Manor Rd, Glasgow	10-Oct-12	1-Dec-13	375	CO93	Tony Shaw
CR56	PG16	Aline Stewart	5 Novar Dr, Glasgow	1-Nov-14	10-Aug-15	450	CO93	Tony Shaw

Client

clientNo	cName
CR76	John Kay
CR56	Aline Stewart

Rental 主键为cn+pn

clientNo	propertyNo	rentStart	rentFinish
CR76	PG4	1-Jul-12	31-Aug-13
CR76	PG16	1-Sep-13	1-Sep-14
CR56	PG4	1-Sep-11	10-Jun-12
CR56	PG36	10-Oct-12	1-Dec-13
CR56	PG16	1-Nov-14	10-Aug-15

STEPS :

- 找出数据表中的所有候选键
- 找出所有的主属性和非主属性
- 判断所有非主属性对候选键的部分函数依赖



大概的意思是一个表只能有一个主键

PropertyOwner

propertyNo	pAddress	rent	ownerNo	oName
PG4	6 Lawrence St, Glasgow	350	CO40	Tina Murphy
PG16	5 Novar Dr, Glasgow	450	CO93	Tony Shaw
PG36	2 Manor Rd, Glasgow	375	CO93	Tony Shaw

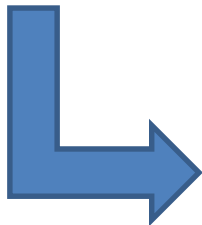
Third Normal Form (3NF, 第三范式)

- A relation that is in second normal form and in which no non-primary-key attribute is 消除传递函数依赖 transitively dependent on the primary key.

PropertyOwner

propertyNo	pAddress	rent	ownerNo	oName
PG4	6 Lawrence St, Glasgow	350	CO40	Tina Murphy
PG16	5 Novar Dr, Glasgow	450	CO93	Tony Shaw
PG36	2 Manor Rd, Glasgow	375	CO93	Tony Shaw

pnum决定onum, onum决定oname
所以oname传递依赖于pnum
拆分为两个表消除传递依赖关系



PropertyForRent

propertyNo	pAddress	rent	ownerNo
PG4	6 Lawrence St, Glasgow	350	CO40
PG16	5 Novar Dr, Glasgow	450	CO93
PG36	2 Manor Rd, Glasgow	375	CO93

Owner

ownerNo	oName
CO40	Tina Murphy
CO93	Tony Shaw

Conversion to First Normal Form (1NF)

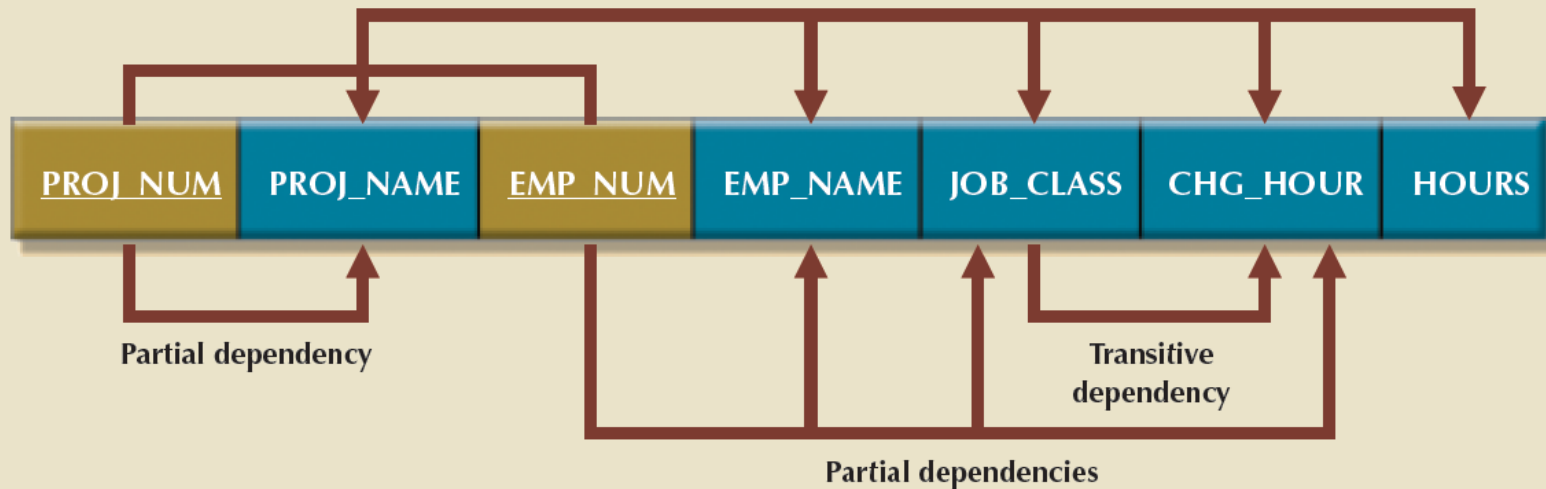
- Repeating group: group of multiple entries of same type can exist for any single key attribute occurrence
 - Reduces data redundancies
- Three step procedure
 - Eliminate the repeating groups
 - Identify the primary key
 - Identify all dependencies
- **Dependency diagram**: depicts all dependencies found within given table structure 依赖图
 - Helps to get an overview of all relationships among table's attributes
 - Makes it less likely that an important dependency will be overlooked

Conversion to First Normal Form (1NF)

- 1NF describes tabular format in which:
 - All key attributes are defined
 - There are no repeating groups in the table
 - All attributes are dependent on the primary key
- All relational tables satisfy 1NF requirements
- Some tables contain partial dependencies
 - Update, insertion, or deletion

Conversion to First Normal Form (1NF)

FIGURE 6.3 FIRST NORMAL FORM (1NF) DEPENDENCY DIAGRAM



1NF (PROJ_NUM, EMP_NUM, PROJ_NAME, EMP_NAME, JOB_CLASS, CHG_HOURS, HOURS)

PARTIAL DEPENDENCIES:

(PROJ_NUM \Rightarrow PROJ_NAME)

(EMP_NUM \Rightarrow EMP_NAME, JOB_CLASS, CHG_HOUR)

TRANSITIVE DEPENDENCY:

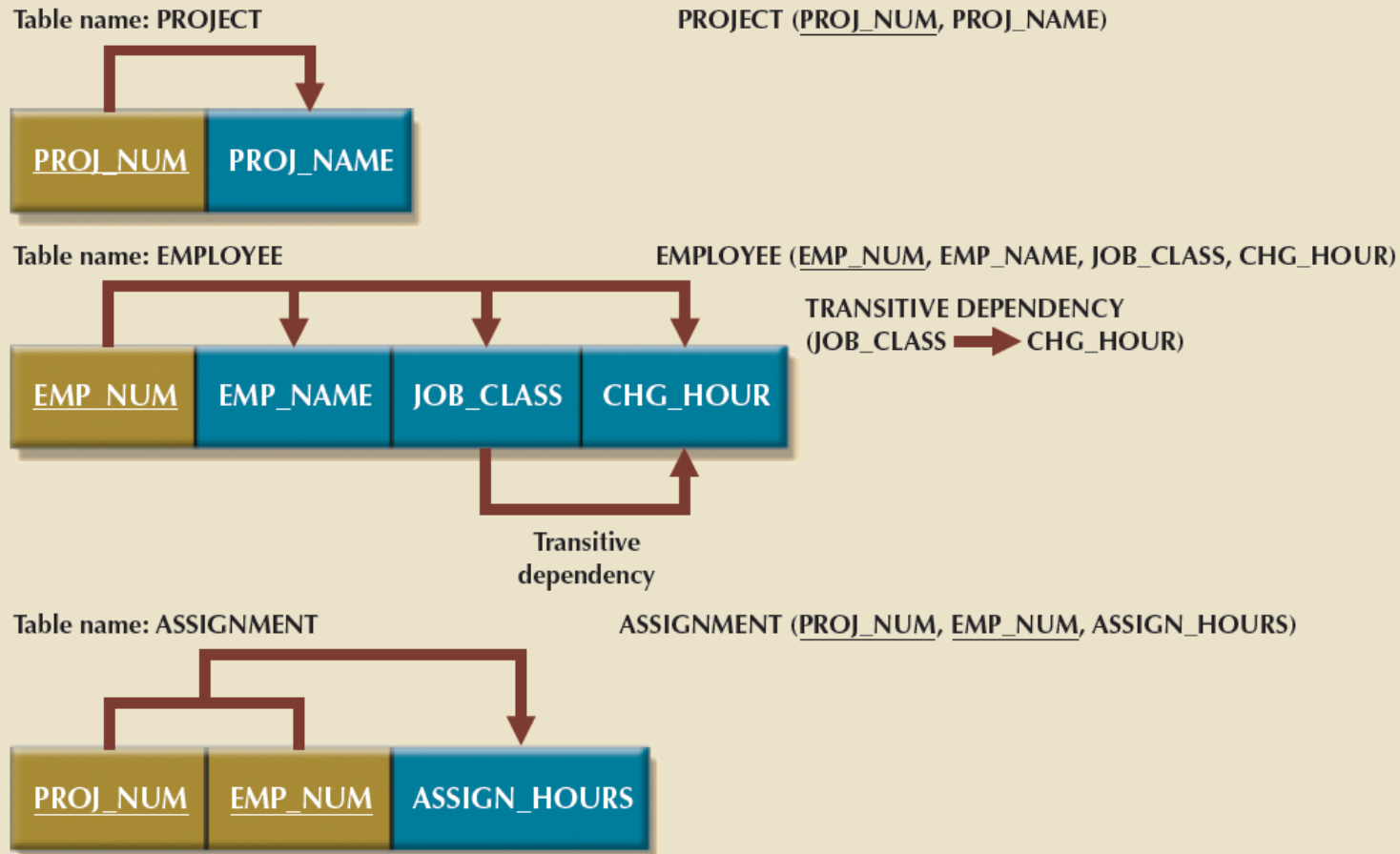
(JOB_CLASS \Rightarrow CHG_HOUR)

Conversion to Second Normal Form (2NF)

- Conversion to 2NF occurs only when the 1NF has a **composite primary key**
 - If the 1NF has a single-attribute primary key, then the table is automatically in 2NF
- The 1NF-to-2NF conversion is simple
 - Make new tables to eliminate partial dependencies
 - Reassign corresponding dependent attributes
- Table is in 2NF when it:
 - Is in 1NF
 - Includes no partial dependencies

Conversion to Second Normal Form (2NF)

FIGURE 6.4 SECOND NORMAL FORM (2NF) CONVERSION RESULTS



Conversion to Third Normal Form (3NF)

- The data anomalies created by the database organization shown in Figure 6.4 are easily eliminated
 - Make new tables to eliminate transitive dependencies
 - Reassign corresponding dependent attributes
- Table is in 3NF when it:
 - Is in 2NF
 - Contains no transitive dependencies

Conversion to Third Normal Form (3NF)

FIGURE 6.5 THIRD NORMAL FORM (3NF) CONVERSION RESULTS



Table name: PROJECT

PROJECT (PROJ_NUM, PROJ_NAME)

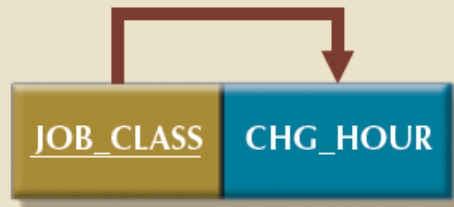


Table name: JOB

JOB (JOB_CLASS, CHG_HOUR)

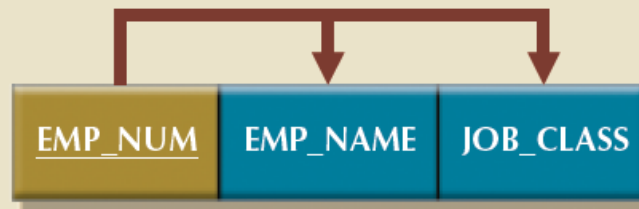


Table name: EMPLOYEE

EMPLOYEE (EMP_NUM, EMP_NAME, JOB_CLASS)

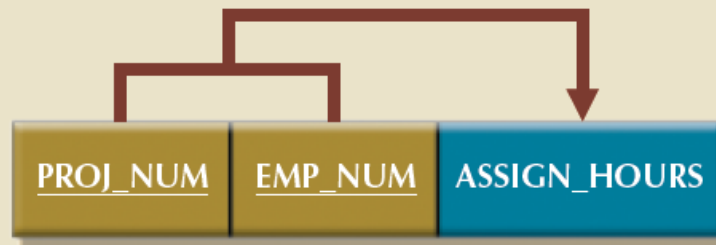


Table name: ASSIGNMENT

ASSIGNMENT (PROJ_NUM, EMP_NUM, ASSIGN_HOURS)

设有一教学管理数据模式，包含如下属性：

学号(S#)，课程号(C#)，成绩(G)，任课教师(TN)，教师所在的系(D)。

这些数据有下列语义：

- ✓ 学号和课程号分别与其代表的学生和课程一一对应。
- ✓ 一个学生所修的每门课程都有一个成绩。
- ✓ 每门课程只有一位任课教师，但每位教师可以有多门课程。
- ✓ 教师中没有重名，每个教师只属于一个系。

(1) 试根据上述语义确定函数依赖集。

(2) 该关系模式为何范式？

(3) 请分解至3NF。

- $R(U, F)$
- $F = \{(S\#, C\#) \rightarrow G, C\# \rightarrow TN, TN \rightarrow D\}$
- 关系模式为1NF
- $R1 = (S\#, C\#, G)$
- $R2 = (C\#, TN)$
- $R3 = (TN, D)$

Conclusions

- Normalization
 - Motivation
 - Concepts
 - FD
 - 1NF
 - 2NF
 - 3NF
 - BCNF
 - Process
 - Conversion
 - Example

Homework

- Read the following Chapters of DS1

- § 6.1-6.5 (pp 200-218)

- § 6.7 (pp 224-227)

- § 9.6 (pp 468-471)

- Assignment

- Later in Yuketang

- Further Reading

- § 5.4 of DS1

- § 14, § 17 of DS2

Thank you!

