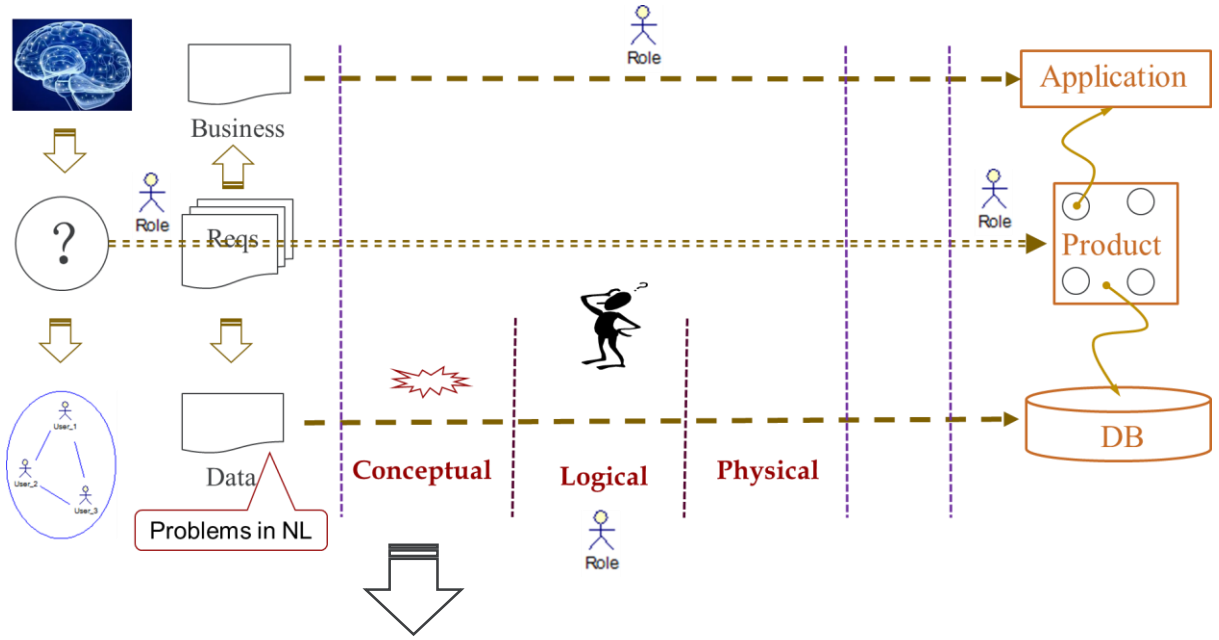
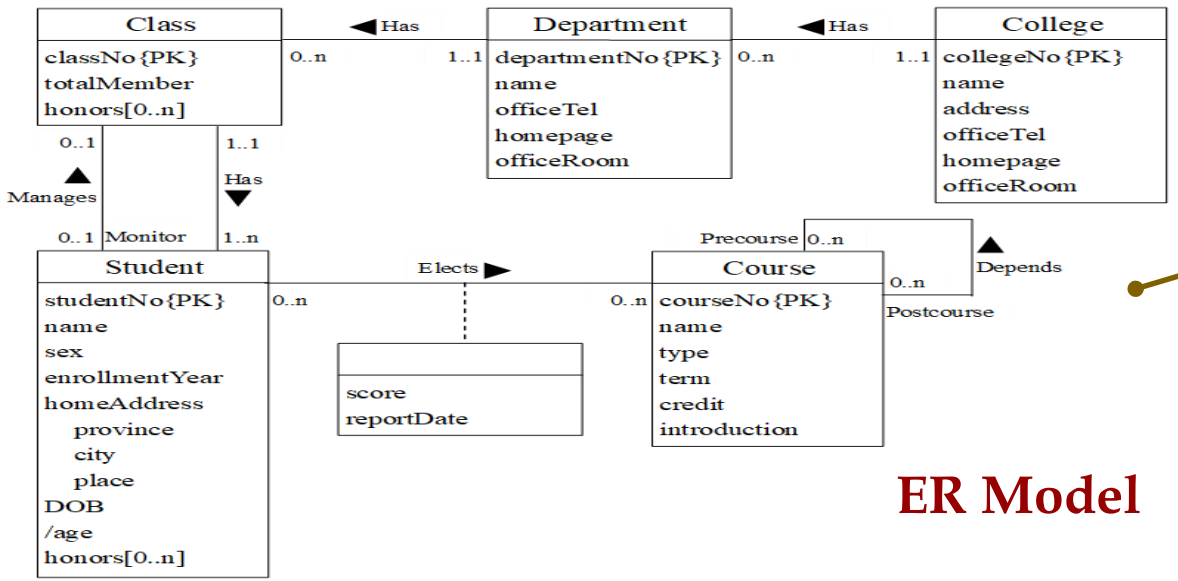
# 关系模式规范化理论

讲解人：陆伟 教授

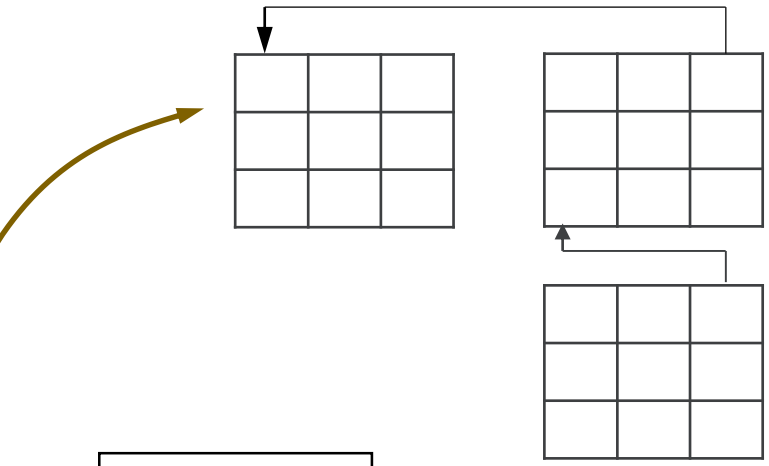
# 规范化问题的引出

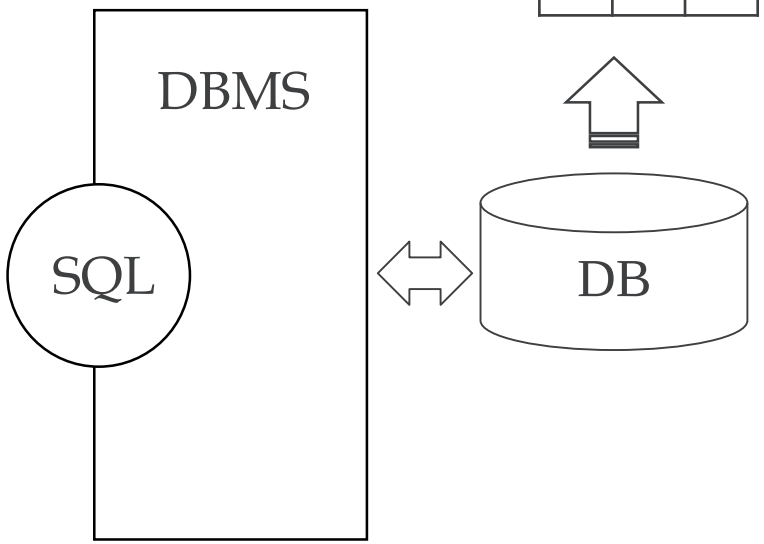



ER Model

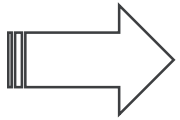
Relation Model





# 规范化问题的引出

| Elements in ER Model | Elements in Relation Model |
| --- | --- |
| entity | relation(table) |
| strong, weak | |
| attribute | attribute(column) |
| simple, composite, mutli-value | |
| relationship | foreign key |
| 1:1, 1:N, M:N | |
| N-ary | |
| recursive | |
| value set | domain |
| key attribute | primary key (alternate key) |



# 规范化问题的引出


D-S(dNo, dName, officeRoom, homepage, sNo, sName, sex, email, age)

D-1(dNo), D-2(dName), D-3(officeRoom), …
S-1(sNo), S-2(sName), S-3(sex), …

D(dNo, dName, officeRoom, homepage, sNo)
S(sNo, sName, sex, email, age)

D(dNo, dName, officeRoom, homepage)
S(sNo, sName, sex, email, age, dNo)

# 关系模式中可能存在的潜在问题分析

| dNo | dName | homepage | office | address |
|---|---|---|---|---|
| 01 | Software | … | … | … |
| 02 | Math | … | … | … |
| … | … | … | … | … |

| tNo | tName | addr | cNo | cName | credit |
|---|---|---|---|---|---|
| T1 | Wangbin | A1 | C1 | N1 | 2 |
| T1 | Wangbin | A1 | C2 | N2 | 2 |
| T1 | Wangbin | A1 | C3 | N3 | 2 |
| T2 | Lina | A2 | C4 | N4 | 3 |
| T2 | Lina | A2 | C5 | N5 | 3 |
| T3 | Zhaole | A3 | C6 | N6 | 2 |

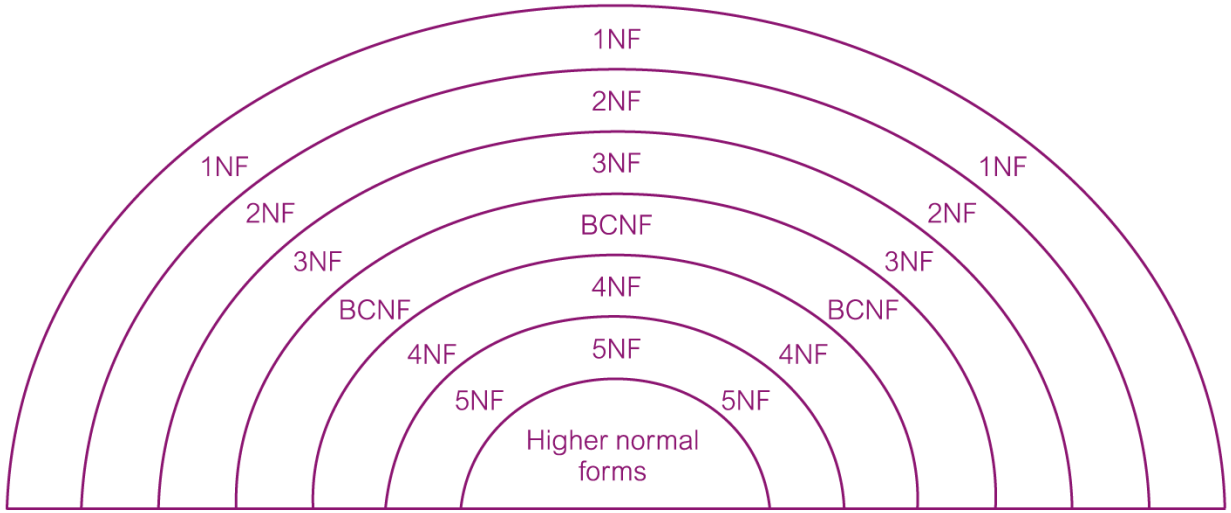
关系模式中可能存在的潜在问题有哪些?

为何有的关系模式不存在这些问题，而有的关系模式存在这些问题?

如何解决这一问题?

# 关系模式规范化概念

- Normalization
  - A technique for producing a set of relations with desirable properties, given the data requirements of an enterprise.
  - The process of normalization is a formal method that identifies relations based on their primary or candidate keys and the data dependencies among their attributes.
- The process of normalization was first developed by E.F.Codd.



# 函数依赖(Functional Dependencies)

- Data dependency
  - The relationships between attributes in a relation that descript the dependency or restriction on each other.
- The types of data dependency
  - Functional dependency(FD) 函数依赖
  - Multi-valued denpendency(MVD) 多值依赖
  - Join denpendency(JD) 连接依赖
- Functional dependency
  - If A and B are attributes of relation R, B is functionally dependent on A(denoted A→B), if each value of A is associated with exactly one value of B. A is determinant of B.

  如何发现关系模式中属性之间的函数依赖关系?

# 函数依赖(Functional Dependencies)

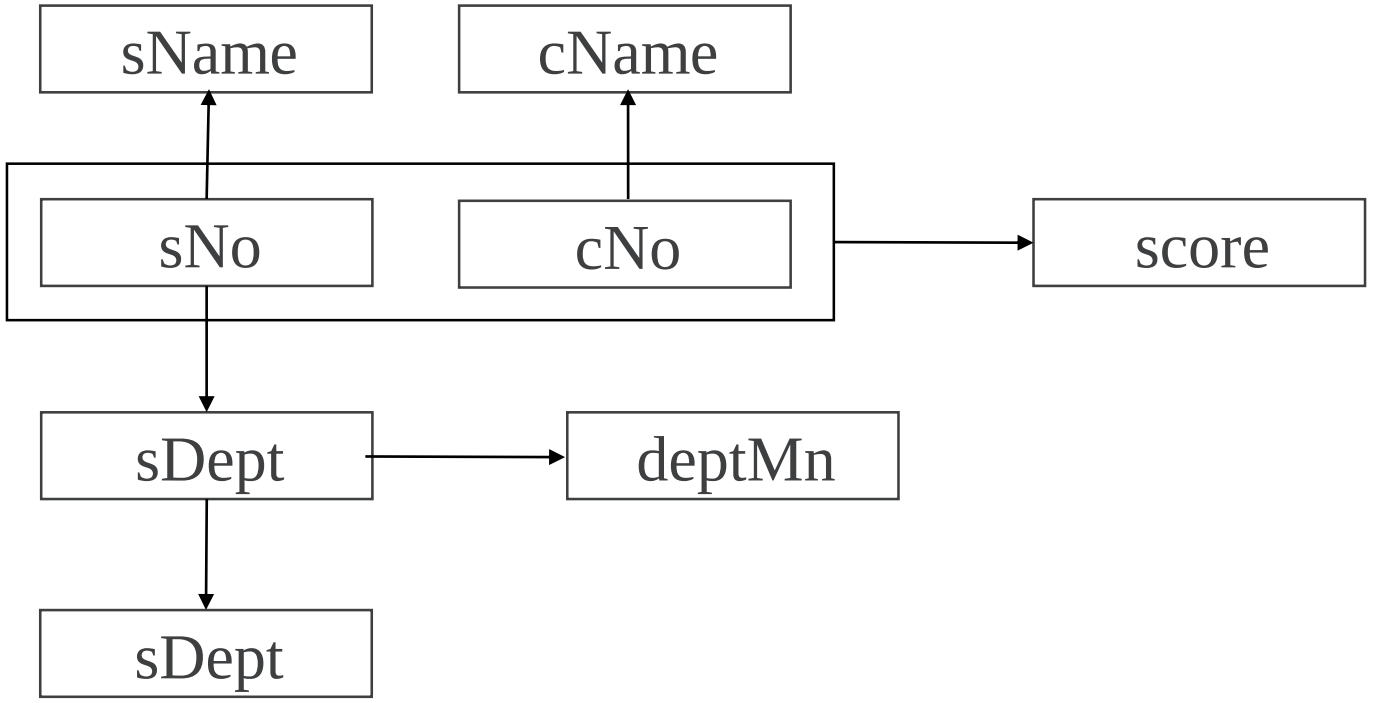
- Functional dependency(FD)
  - If $A \rightarrow B$ and $B \rightarrow A$, then we denote $A \leftrightarrow B$
  - Trivial functional dependency(平凡函数依赖)：If $A \rightarrow B$ and B is a subset of A($B \subseteq A$)
  - Nontrivial functional dependency：If $A \rightarrow B$ and B is not a subset of A($B \subseteq A$)
- Generally we can identify:
  - If the relationship between X and Y is one-to-one(1:1), then $X \leftrightarrow Y$
  - If the relationship between X and Y is one-to-many(1:*), then $Y \rightarrow X$
  - If the relationship between X and Y is many-to-many(*:*), then maybe there is no functional dependency between X and Y

# 函数依赖(Functional Dependencies)

关系模式R(U, F)

U={sNo, sName, sDept, deptAddr, deptMn, cNo, cName, score}

F={sNo→sName; sNo→sDept; sDept→deptAddr,deptMn;
cNo→cName; (sNo,cNo)→score}



# 函数依赖(Functional Dependencies)

- The set of all functional dependencies that are implied by a given set of functional dependencies X is called the closure of X, written $X^+$. Minimal Sets of Functional Dependencies($X_{min}$)
- A set of inference rules, called Armstrong's axioms, can help to compute $X^+$ from X.
- Armstrong's axioms are as follows:

  Reflexivity: If B is a subset of A, then $A \rightarrow B$

  Augmentation: If $A \rightarrow B$, then $A,C \rightarrow B,C$

  Transitivity: If $A \rightarrow B$ and $B \rightarrow C$, then $A \rightarrow C$.

  Self-determination: $A \rightarrow A$

  Decomposition: If $A \rightarrow B,C$, then $A \rightarrow B$ and $A \rightarrow C$.

  Union: If $A \rightarrow B$ and $A \rightarrow C$ , then $A \rightarrow B,C$

  Composition: If $A \rightarrow B$ and $C \rightarrow D$ , then $A,C \rightarrow B,D$

# 第一范式(First Normal Form, 1NF)

- Definition of 1NF
  - A relation in which the intersection of each row and column contains one and only one value.
  - 1NF是关系模式规范化的最低要求

| sNo | sName | addr | phoneNo |
|---|---|---|---|
| s01 | 赵剑 | No.1 | 02988451234<br>02886654321 |
| s02 | 钱斌 | No.2 | 02988451235 |
| s03 | 孙兰 | No.3 | 02988451236<br>02886654322 |

该存储模式是否满足1NF？如何改造，使其满足1NF?

# 第一范式(First Normal Form, 1NF)

| sNo | sName | addr | phoneNo |
|---|---|---|---|
| s01 | 赵剑 | No.1 | 02988451234 |
| s01 | 赵剑 | No.1 | 02886654321 |
| s02 | 钱斌 | No.2 | 02988451235 |
| s03 | 孙兰 | No.3 | 02988451236 |
| s03 | 孙兰 | No.3 | 02886654322 |

| sNo | sName | addr | phoneNo1 | phoneNo2 |
|---|---|---|---|---|
| s01 | 赵剑 | No.1 | 02988451234 | 02886654321 |
| s02 | 钱斌 | No.2 | 02988451235 | |
| s03 | 孙兰 | No.3 | 02988451236 | 02886654322 |

讨论：两种改造方式的各自优缺点是什么?

# 第二范式(Second Normal Form, 2NF)

- Problems in a relation that is in 1NF

| sNo | cNo | score | credit |
|---|---|---|---|
| s01 | c01 | 80 | 2 |
| s01 | c02 | 90 | 3 |
| s02 | c01 | 70 | 2 |
| s02 | c03 | 85 | 1 |
| s03 | c01 | 75 | 2 |
| s04 | c02 | 85 | 3 |

分析：该关系模式满足1NF，是否还存在问题?

讨论：该关系模式存在问题的原因

# 第二范式(Second Normal Form, 2NF)

## Full/Partial Functional Dependency

- $X \rightarrow Y$ is a FD in R(U,F), if $\neg \exists X' \subset X$, $X' \rightarrow Y$, then Y is fully functionally dependent on $X$. Or $Y$ is partially functionally dependent on $X$.
- We can write Partial/Full functional dependency as

$$X \xrightarrow{P} Y \qquad X \xrightarrow{F} Y$$

## Definition of 2NF

- A relation that is in 1NF and every non-primary-key attribute is fully functionally dependent on the primary key.
- 第二范式要求关系模式中不能存在非主码属性对主码的部分依赖。

问题：如果一个关系模式主码为单个属性，该关系模式是否一定满足2NF?

# 第二范式(Second Normal Form, 2NF)

- The normalization of 1NF relations to 2NF involves the removal of partial dependencies.
  - If a partial dependency exists, we remove the partially functionally dependent attribute(s) from the relation by placing them in a new relation along with a copy of their determinant.
- Example

关系模式R(U, F)

U={sNo, cNo, score, credit}

F={ (sNo,cNo)→score; cNo→credit}

| sNo | cNo | score | credit |
|---|---|---|---|
| s01 | c01 | 80 | 2 |
| s01 | c02 | 90 | 3 |
| s02 | c01 | 70 | 2 |
| s02 | c03 | 85 | 1 |
| s03 | c01 | 75 | 2 |
| s04 | c02 | 85 | 3 |

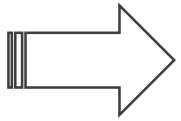
# 第二范式(Second Normal Form, 2NF)

关系模式R(U, F)

U={sNo, cNo, score, credit}

F={ (sNo,cNo)→score; cNo→credit}

| sNo | cNo | score | credit |
|---|---|---|---|
| s01 | c01 | 80 | 2 |
| s01 | c02 | 90 | 3 |
| s02 | c01 | 70 | 2 |
| s02 | c03 | 85 | 1 |
| s03 | c01 | 75 | 2 |
| s04 | c02 | 85 | 3 |



We can decompose R to R1 and R2

R1: U={sNo, cNo, score}  F={(sNo,cNo)→score}

R2: U={cNo, credit}  F={ cNo→credit}

| sNo | cNo | score |
|---|---|---|
| s01 | c01 | 80 |
| s01 | c02 | 90 |
| s02 | c01 | 70 |
| s02 | c03 | 85 |
| s03 | c01 | 75 |
| s04 | c02 | 85 |

| cNo | credit |
|---|---|
| c01 | 2 |
| c02 | 3 |
| c03 | 1 |

# 第三范式(Third Normal Form, 3NF)

- Problems in a relation that is in 2NF

| sNo | sName | dept | deptAddr | deptMn |
|---|---|---|---|---|
| s01 | 赵剑 | CS | Building 1 | 赵彬 |
| s02 | 钱斌 | IS | Building 2 | 孙淼 |
| s03 | 孙兰 | Math | Building 3 | 钱思 |
| s04 | 李锋 | CS | Building 1 | 赵彬 |
| s05 | 王谦 | IS | Building 2 | 孙淼 |
| … | … | … | … | … |

分析：该关系模式满足2NF，是否还存在问题?

讨论：该关系模式存在问题的原因

# 第三范式(Third Normal Form, 3NF)

- Transitive Dependency
  - A condition where A, B, and C are attributes of a relation such that if $A \rightarrow B$ and $B \rightarrow C (C \not\subseteq B)$, then C is transitively dependent on A via B(provided that A is not functionally dependent on B or C).
- The normalization of 2NF relations to 3NF involves the removal of transitive dependencies.
  - If a transitive dependency exists, we remove the transitively dependent attribute(s) from the relation by placing them in a new relation along with a copy of their determinant.

# 第三范式(Third Normal Form, 3NF)

## Example

关系模式R(U, F)

U={sNo, sName, dept, deptAddr, deptMn}

F={ sNo→sName, dept; dept→deptAddr, deptMn}

| sNo | sName | dept | deptAddr | deptMn |
|---|---|---|---|---|
| s01 | 赵剑 | CS | Building 1 | 赵彬 |
| s02 | 钱斌 | IS | Building 2 | 孙淼 |
| s03 | 孙兰 | Math | Building 3 | 钱思 |
| s04 | 李锋 | CS | Building 1 | 赵彬 |
| s05 | 王谦 | IS | Building 2 | 孙淼 |
| … | … | … | … | … |

We can decompose R to R1 and R2

R1: U={sNo, sName, dept}

F={sNo→sName, dept}

R2: U={dept, deptAddr, deptMn}

F={dept→deptAddr, deptMn}

| sNo | sName | dept |
|---|---|---|
| s01 | 赵剑 | CS |
| s02 | 钱斌 | IS |
| s03 | 孙兰 | Math |
| s04 | 李锋 | CS |
| s05 | 王谦 | IS |

| dept | deptAddr | deptMn |
|---|---|---|
| CS | Building 1 | 赵彬 |
| IS | Building 2 | 孙淼 |
| Math | Building 3 | 钱思 |

# 1NF到3NF过程回顾

| sNo | sName | dept | deptAddr | deptMn | cNo | cName | credit | score |
|---|---|---|---|---|---|---|---|---|
| s01 | 赵剑 | CS | Building 1 | 赵彬 | c01 | DB | 2 | 80 |
| s01 | 赵剑 | CS | Building 1 | 赵彬 | c02 | OS | 3 | 90 |
| s02 | 钱斌 | IS | Building 2 | 孙淼 | c01 | DB | 2 | 70 |
| s02 | 钱斌 | IS | Building 2 | 孙淼 | c03 | DS | 1 | 85 |
| s03 | 孙兰 | Math | Building 3 | 钱思 | c01 | DB | 2 | 75 |
| s04 | 李锋 | CS | Building 1 | 赵彬 | c02 | OS | 3 | 85 |

关系模式StuDeptCourse(U, F)

U={sNo, sName, dept, deptAddr, deptMn, cNo, cName, credit, score}

F={ sNo→sName,dept; dept→deptAddr,deptMn; cNo→cName, credit; (sNo,cNo)→score}

# 1NF到3NF回顾

| sNo | cNo | score |
|---|---|---|
| s01 | c01 | 80 |
| s01 | c02 | 90 |
| s02 | c01 | 70 |
| s02 | c03 | 85 |
| s03 | c01 | 75 |
| s04 | c02 | 85 |

| sNo | sName | dept | deptAddr | deptMn |
|---|---|---|---|---|
| s01 | 赵剑 | CS | Building 1 | 赵彬 |
| s02 | 钱斌 | IS | Building 2 | 孙淼 |
| s03 | 孙兰 | Math | Building 3 | 钱思 |
| s04 | 李锋 | CS | Building 1 | 赵彬 |

| cNo | cName | credit |
|---|---|---|
| c01 | DB | 2 |
| c02 | OS | 3 |
| c03 | DS | 1 |

关系模式SC(U, F)
U={sNo, cNo, score}
F={(sNo,cNo)→score}

SC∈2NF, SC∈3NF

关系模式SD(U, F)
U={sNo, sName, dept, deptAddr, deptMn}
F={sNo→sName, dept, deptAddr, deptMn;
dept →deptMn}

SD∈2NF, SC∉3NF

关系模式C(U, F)
U={cNo, cName, credit}
F={cNo→cName,credit}

C∈2NF, SC∈3NF

# 1NF到3NF过程回顾

| sNo | cNo | score |
|---|---|---|
| s01 | c01 | 80 |
| s01 | c02 | 90 |
| s02 | c01 | 70 |
| s02 | c03 | 85 |
| s03 | c01 | 75 |
| s04 | c02 | 85 |

| cNo | cName | credit |
|---|---|---|
| c01 | DB | 2 |
| c02 | OS | 3 |
| c03 | DS | 1 |

| sNo | sName | dept |
|---|---|---|
| s01 | 赵剑 | CS |
| s02 | 钱斌 | IS |
| s03 | 孙兰 | Math |
| s04 | 李锋 | CS |

| dept | deptAddr | deptMn |
|---|---|---|
| CS | Building 1 | 赵彬 |
| IS | Building 2 | 孙淼 |
| Math | Building 3 | 钱思 |

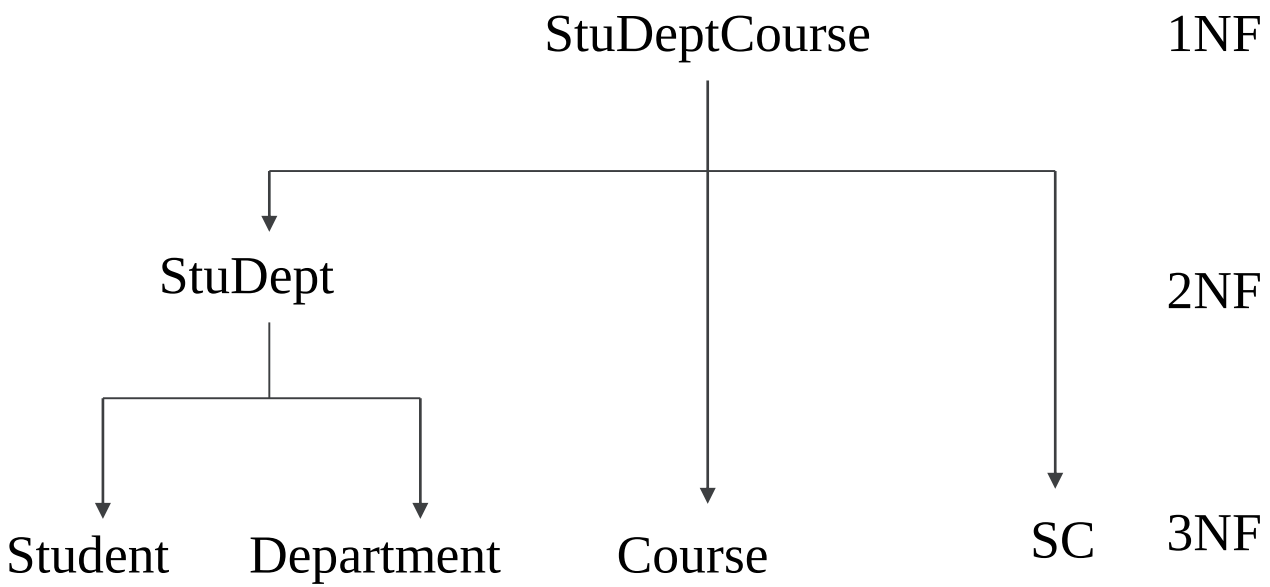
关系模式SC(U, F)
U={sNo, cNo, score}
F={(sNo,cNo)→score}

关系模式C(U, F)
U={cNo, cName, credit}
F={cNo→cName,credit}

关系模式S(U, F)
U={sNo, sName, dept}
F={sNo→sName,dept}

关系模式D(U, F)
U={dept, deptAddr, deptMn}
F={dept→deptAddr,deptMn}

# 1NF到3NF过程回顾



# 2NF和3NF一般化定义

问题：若一个关系模式具有2个或者2个以上的候选码，非码属性对所有候选码的部分依赖和传递依赖是否会导致同样的问题?

- Definition of 2NF in general
  - A relation that is in first normal form and every non-primary-key attribute is fully functionally dependent on any candidate key.
- Definition of 3NF in general
  - A relation that is in first and second normal form and in which no non-primary-key attribute is transitively dependent on any candidate key.

# Boyce-Codd Noraml Form (BCNF)

- 部分依赖和传递依赖是导致关系模式中存在数据更新异常的主要原因，第三范式通过消除这些部分依赖和传递依赖，进而能够获得良好的结果。
- 在数据库设计中，如果所有关系模式都满足第三范式要求，我们可以认为这是一个不错的结果。

问题：第三范式并没有考虑主属性本身对码的部分依赖和传递依赖，这些主属性本身对码的部分依赖和传递依赖是否仍然可能导致问题?

关系模式Order(U, F)－餐馆点菜系统

U={roomId, foodId, waiterId, cookId}

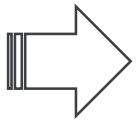
F={(roomId, foodId)→waiterId, cookId; (foodId, waiterId)→cookId; waiterId→roomId}

$Order \in 3NF, Order \notin BCNF$

# Boyce-Codd Noraml Form (BCNF)

- Definnition of BCNF
  - A relation is in BCNF, if and only if, every determinant is a candidate key.
  - Or no attribute in a relation is transitively dependent on any candidate key.

| roomId | foodId | waiterId | cookId |
|---|---|---|---|
| 1 | 1 | 1 | 1 |
| 1 | 2 | 1 | 2 |
| 1 | 3 | 2 | 2 |
| 2 | 1 | 3 | 3 |
| 2 | 3 | 3 | 2 |
| 2 | 4 | 4 | 1 |



| foodId | waiterId | cookId |
|---|---|---|
| 1 | 1 | 1 |
| 2 | 1 | 2 |
| 3 | 2 | 2 |
| 1 | 3 | 3 |
| 3 | 3 | 2 |
| 4 | 4 | 1 |

| waiterId | roomId |
|---|---|
| 1 | 1 |
| 2 | 1 |
| 3 | 2 |
| 4 | 2 |

讨论：一般什么情况下会出现一个关系模式满足3NF而不满足BCNF?

# 随堂测试

给定关系模式R(U, F)，

U={A, B, C, D, E, F, G}，

F={A→B,C; D→E; AD→F; E→G}。

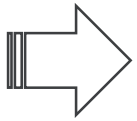
（1）求该关系模式候选码

（2）将关系模式R一步一步规范化为满足BCNF的关系模式。

# 第四范式(Fourth Normal Form, 4NF)

- Although BCNF removes anomalies due to functional dependencies, another type of dependency called a multi-valued dependency (MVD) can also cause data redundancy.

| course | teacher | references |
|---|---|---|
| math | 孙丽<br>李苗 | 高等数学<br>数学分析 |
| physics | 赵明<br>钱丰 | 大学物理<br>物理习题集<br>理论力学 |



| course | teacher | references |
|---|---|---|
| math | 孙丽 | 高等数学 |
| math | 孙丽 | 数学分析 |
| math | 李苗 | 高等数学 |
| math | 李苗 | 数学分析 |
| physics | 赵明 | 大学物理 |
| physics | 赵明 | 物理习题集 |
| physics | 赵明 | 理论力学 |
| physics | 钱丰 | 大学物理 |
| … | … | … |

# 第四范式(Fourth Normal Form, 4NF)

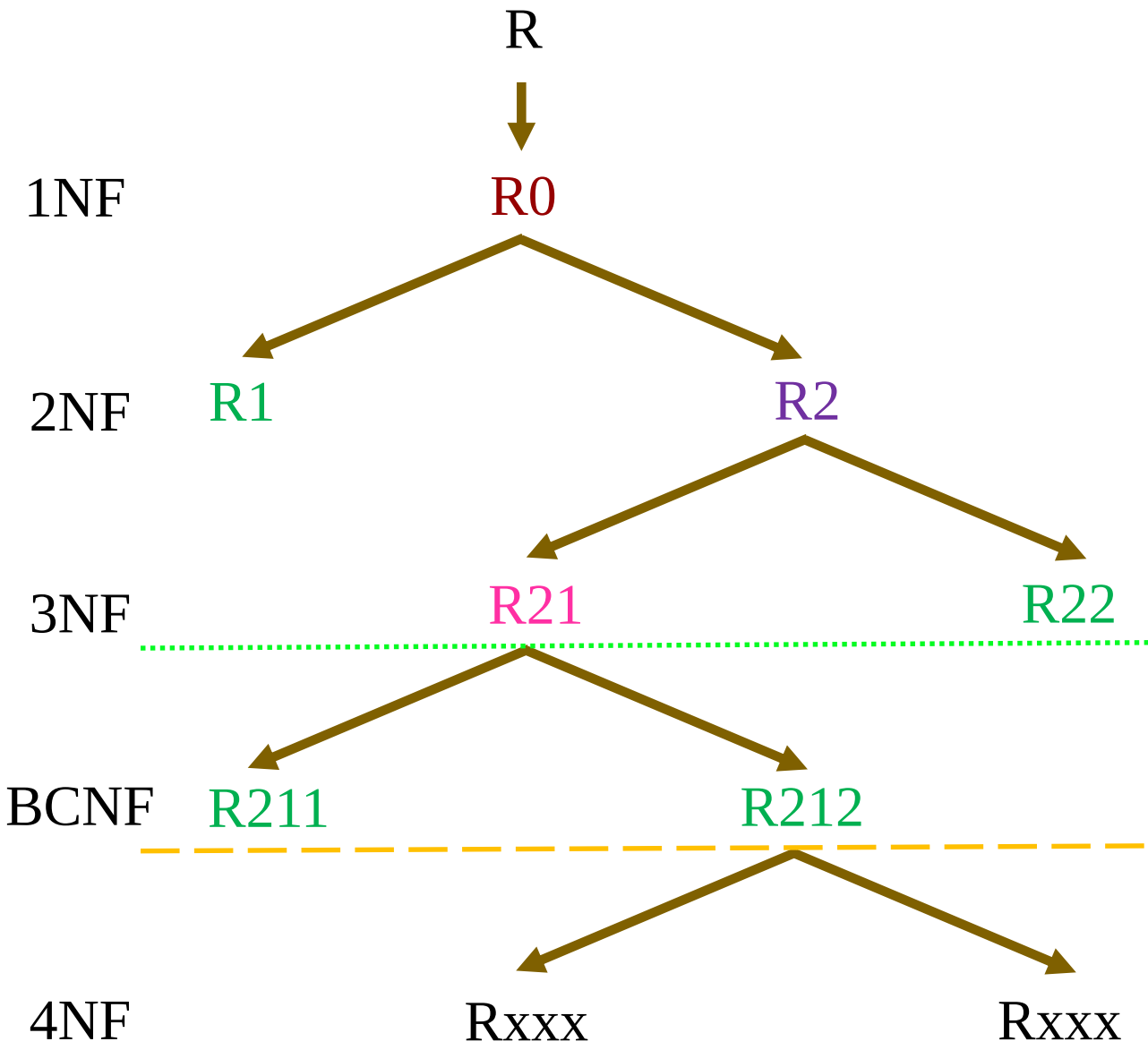
- Multi-valued dependency (MVD)
  - Dependency between attributes (for example, A, B, and C) in a relation, such that for each value of A there is a set of values for B and a set of values for C. However, the set of values for B and C are independent of each other.

$$A \rightarrow\rightarrow B$$

$$A \rightarrow\rightarrow C$$

- Definition of 4NF
  - A relation that is in BCNF and contains no nontrivial MVDs.
  - The normalization of BCNF relations to 4NF involves the removal of the MVD from the relation by placing the attribute(s) in a new relation along with a copy of the determinant(s).

# 关系模式规范化的工程思考



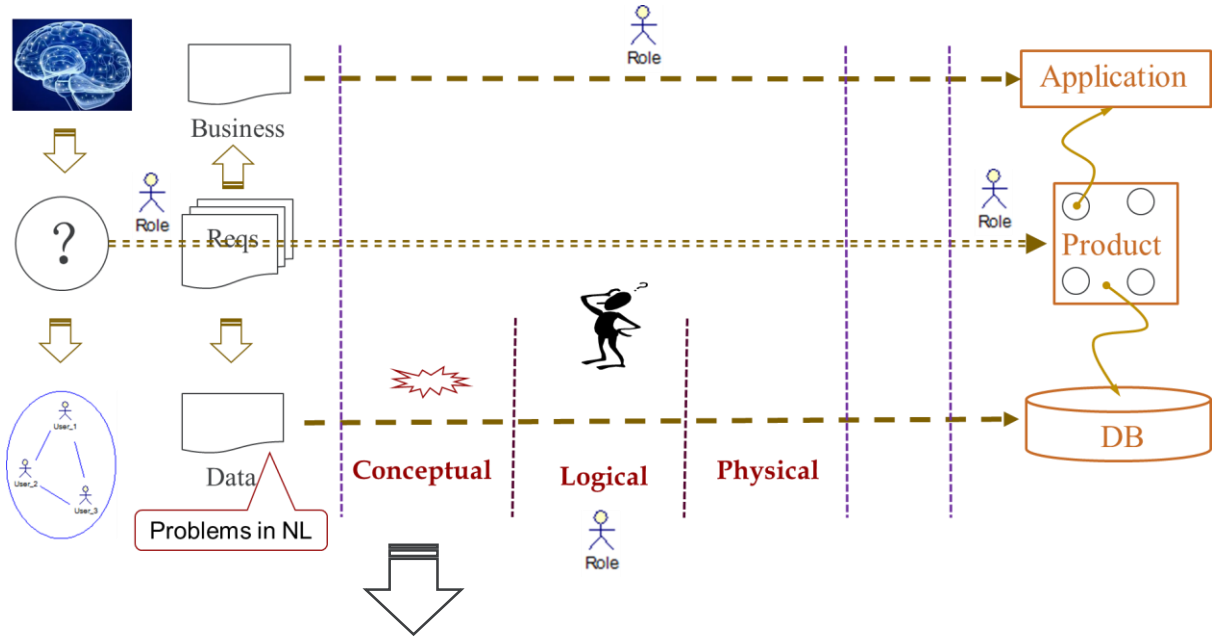

讨论：关系模式规范化仅仅是理论，并不是标准，我们在实际工程中应该如何做？

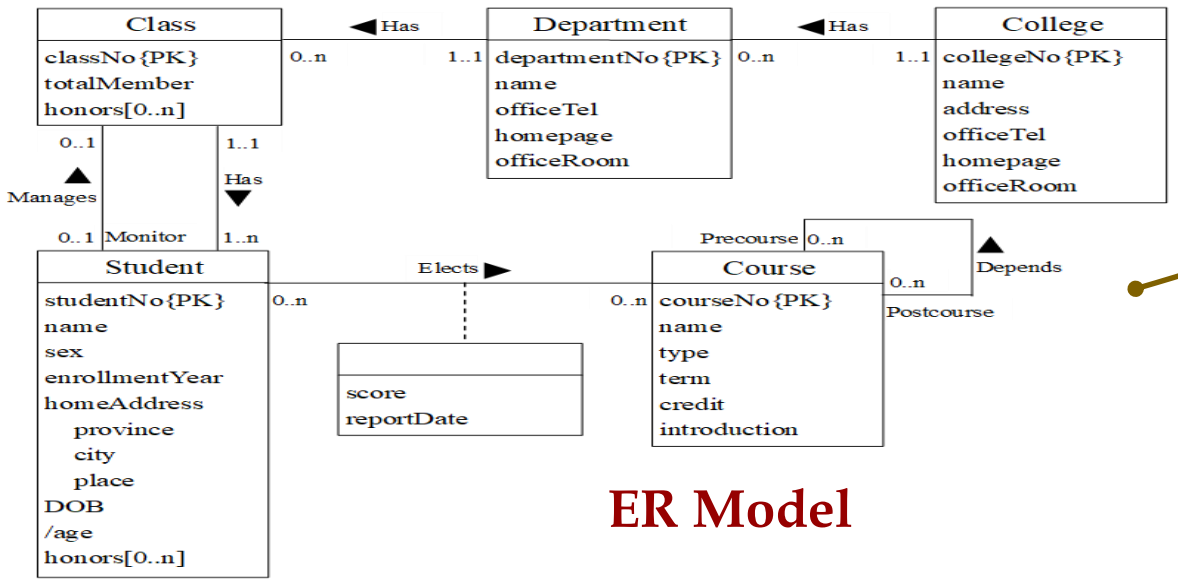
消除了大部分数据更新异常和数据冗余，并且能够保持原来关系模式中的所有函数依赖

在函数依赖的范围内实现了彻底分解，消除了所有更新异常问题。

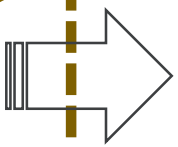
关系模式分解之后效率如何？

# 总结

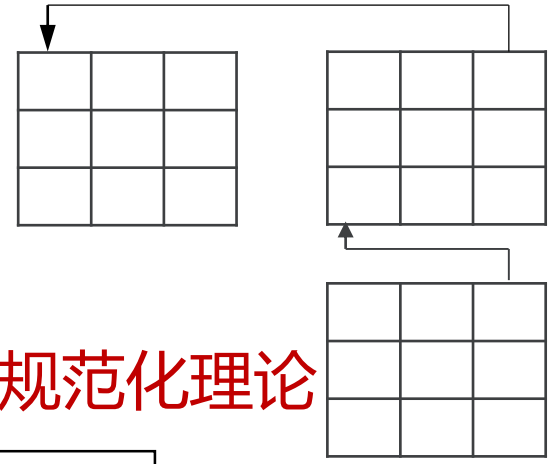


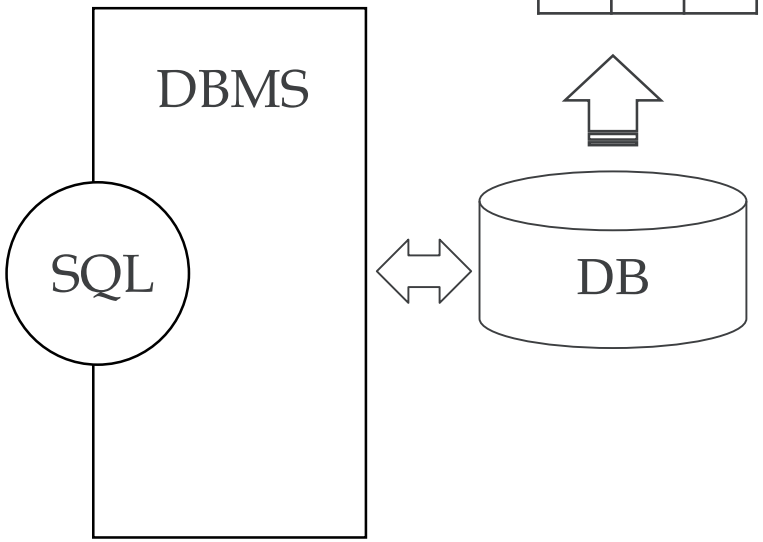

ER Model



Relation Model



关系模式规范化理论



# 关于本讲内容

**祝各位学习愉快！**

# 感谢观看！

讲解人：陆伟 教授