数据库系统之三
--数据建模与数据库设计

第15讲 关系模式设计之规范形式

战德臣

哈尔滨工业大学　教授.博士生导师

黑龙江省教学名师

教育部大学计算机课程教学指导委员会委员

Research Center on Intelligent Computing for Enterprises & Services, Harbin Institute of Technology

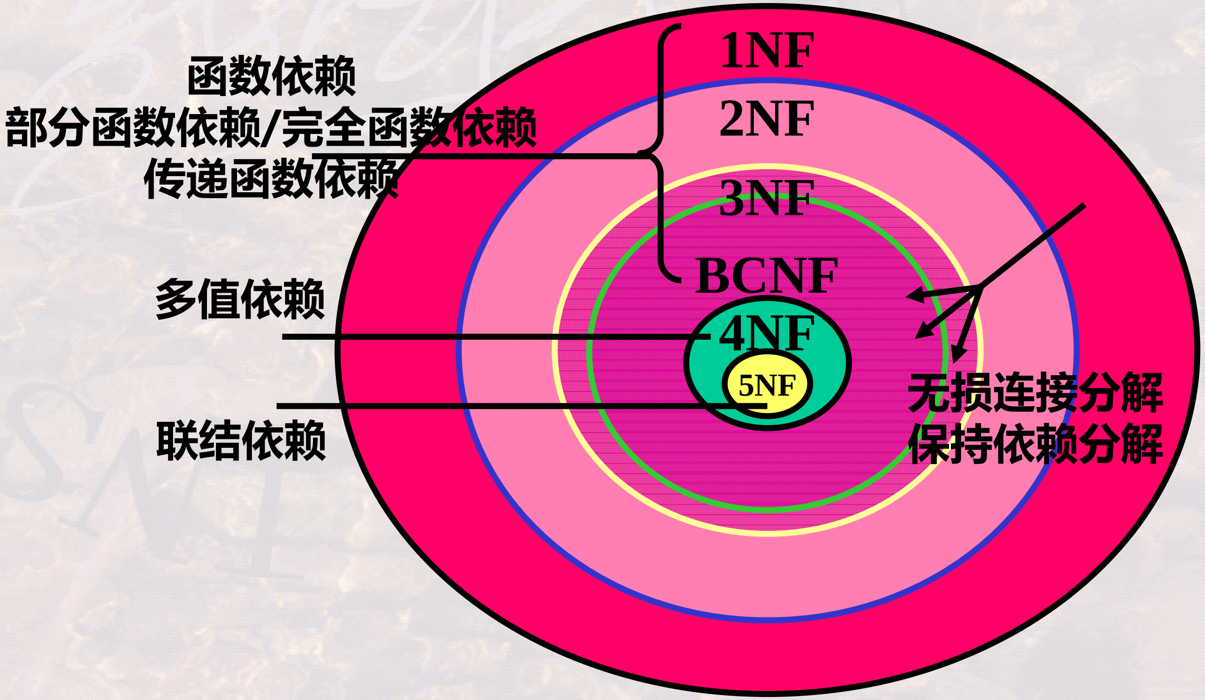
本讲学习什么?

战德臣 教授

如何避免数据库的一致性问题—数据库的规范性设计

数据库的规范性设计需要分析数据库Table中的属性在取值方面有什么依存关系？数据库设计过程中应遵循什么样的原则

- **数据库设计理论**
  - ❑数据依赖理论
  - ❑关系范式理论
  - ❑模式分解理论



本讲学习什么?

战德臣 教授

基本内容

1. 关系的第1NF和第2NF
2. 关系的第3NF和Boyce-Codd NF
3. 多值依赖及其公理定理
4. 关系的第4NF

重点与难点

- 一组概念：1NF, 2NF, 3NF, BCNF, 4NF；多值依赖
- 熟练应用数据库设计的规范化形式，判断数据库设计的正确性及可能存在的问题

关系的第1范式和第2范式

战德臣

哈尔滨工业大学　教授.博士生导师
黑龙江省教学名师
教育部大学计算机课程教学指导委员会委员

Research Center on Intelligent
Computing for Enterprises & Services,
Harbin Institute of Technology

战德臣 教授

[Definition] **1NF**

若关系模式R(U)中关系的每个分量都是不可分的数据项(值、原子)，则称R(U)属于第一范式，记为：R(U)∈1NF。

示例： Star(name, address(street, city))

Star不属于1NF, 因为属性address仍包含了street和city两个属性，其分量不是原子。

Students

sid	lname	fname	class	telephone
1	Jones	Allan	2	555-1234
2	Smith	John	3	555-4321
3	Brown	Harry	2	555-1122
5	White	Edward	3	555-3344

符合1NF

1NF要求关系中不能有复合属性、多值属性及其组合

复合属性

Head: structured type

Students

sid	name		class	telephone	enrollment	
	lname	fname			cno	major
1	Jones	Allan	2	555-1234	101	No
					108	Yes
2	Smith	John	3	555-4321	105	No
3	Borwn	Harry	2	555-1122	101	Yes
					108	No
4	White	Edward	3	555-3344	102	No
					105	No

不符合1NF

多值属性

Value: structured value collection of values

不符合1NF的处理

将 非1NF转换为 1NF情况

示例： Star(name, address(street, city))

➔ Star(name, address) 或者 Star (name, street, city)

将复合属性处理为简单属性；将多值属性与关键字单独组成一新的关系

引入新的数据模型处理：Object-Oriented Data Model

列对象

Head: structured type

Students

sid	name		class	telephone	enrollment	
	lname	fname			cno	major
1	Jones	Allan	2	555-1234	101	No
					108	Yes
2	Smith	John	3	555-4321	105	No
3	Borwn	Harry	2	555-1122	101	Yes
					108	No
4	White	Edward	3	555-3344	102	No
					105	No

行对象

结构对象
聚集对象

Value: structured value
collection of values

关系的第1范式和第2范式
(2)关系的2NF

战德臣 教授

[Definition] 2NF

若R(U)$\in$1NF 且 U中的每一非主属性完全函数依赖于候选键，则称R(U)属于第二范式，记为：R(U)$\in$2NF。

第二范式消除了非主属性对候选键的部分依赖。

示例：R(S#, SN, SD, CN, G)

其中，S#:学号, SN:姓名, SD:班级, CN:课程, G:成绩。

函数依赖：S#$\rightarrow$SN, S#$\rightarrow$SD, {S#, CN}$\rightarrow$G

候选键：{S#,CN}，非主属性：SN和SD。

因为：{S#, CN} $\xrightarrow{p}$ {SN、SD}，所以R不属于2NF。

将其分解为R_1(S#, SN, SD), R_2(S#, CN, G), 则$R_1\in$2NF，$R_2\in$2NF。

学生

学号	姓名	课程号	课程名	成绩
98030101	张三	001	数据库	92
98030101	张三	002	计算机原理	85
98030101	张三	003	高等数学	88
98040202	李四	002	计算机原理	90
98040202	李四	003	高等数学	80
98040202	李四	001	数据库	55
98040203	王五	003	高等数学	56
98030102	周六	001	数据库	54
98030102	周六	002	计算机原理	85
98030102	周六	003	高等数学	48

学生

学号	姓名
98030101	张三
98040202	李四
98040203	王五
98030102	周六

课程

课程号	课程名
001	数据库
002	计算机原理
003	高等数学

选课

学号	课程号	成绩
98030101	001	92
98030101	002	85
98030101	003	88
98040202	002	90
98040202	003	80
98040202	001	55
98040203	003	56
98030102	001	54
98030102	002	85
98030102	003	48

练习：下列模式是否满足第2范式？怎样使其满足第2范式？

- 学生(学号, 姓名, 班级, 课号, 课程名, 成绩, 教师, 教师职务)
  - 候选键 ：{学号，课号} $\xrightarrow{f}$ U；非主属性：姓名、课程名
  - 部分依赖： {学号，课号} $\xrightarrow{p}$ 课程名； {学号，课号} $\xrightarrow{p}$ 姓名
- 员工(员工码, 姓名, 出生日期, 联系电话, 最后学历, 毕业学校, 培训日期, 培训内容)
  - 候选键： {员工码，培训日期} $\xrightarrow{f}$ U；非主属性：姓名，出生日期
  - 部分依赖：{员工码，培训日期} $\xrightarrow{p}$ { 姓名, 出生日期 }；
- 图书(书号, 书名, 出版日期, 出版社, 书架号, 房间号)
  - 候选键： 书号 $\xrightarrow{f}$ U； 非主属性：候选键外其他属性
  - 无部分依赖： 书号 $\xrightarrow{f}$ 每一个属性；

课后思考：举出一些满足第2范式的实例和不满足第2范式的实例

关系的第3范式和Boyce-Codd范式

战德臣

哈尔滨工业大学　教授.博士生导师

黑龙江省教学名师

教育部大学计算机课程教学指导委员会委员

Research Center on Intelligent
Computing for Enterprises & Services,
Harbin Institute of Technology

[Definition]3NF

若R(U，F)$\in$2NF 且 R中不存在这样的情况：候选键X，属性组Y$\subseteq$U和非主属性A, 且A$\notin$X, A$\notin$Y，Y$\not\subset$X, Y$\not\rightarrow$X，使得X$\rightarrow$Y，Y$\rightarrow$A成立。满足以上条件则称R(U)属于第三范式，记为：R(U)$\in$3NF。

示例：**Store(Sid, Pid, Did, Mgr)**

其中，Sid:商店, Pid:商品, Did:经营部, Mgr:经理。

函数依赖：{Sid, Pid} $\rightarrow$ Did, {Sid, Did}$\rightarrow$Mgr

候选键：{Sid,Pid} ，非主属性：Mgr。

因为：{Sid, Pid}→Did, {Sid, Did}→Mgr ，所以R不属于3NF。

将其分解为R_1(Sid, Pid, Did), R_2(Sid, Did, Mgr),

则$R_1\in$3NF，$R_2\in$3NF。

商店

商店	商品(编号)	商品部	商品部经理
一分店	鞋01	鞋类部	张三
一分店	鞋02	鞋类部	张三
一分店	鞋03	鞋类部	张三
一分店	鞋04	鞋类部	张三
一分店	鞋05	鞋类部	张三
一分店	化装品01	化装品部	李四
一分店	化装品02	化装品部	李四
一分店	化装品03	化装品部	李四
一分店	化装品04	化装品部	李四
一分店	化装品05	化装品部	李四
一分店	化装品06	化装品部	李四
二分店	鞋01	综合部	王三
二分店	鞋02	综合部	王三
二分店	鞋03	综合部	王三
二分店	化装品01	妇儿部	王四
二分店	化装品02	妇儿部	王四
二分店	化装品03	妇儿部	王四

➢第3范式消除了非主属性对侯选键的传递依赖

练习：下列模式是否满足第3范式？怎样使其满足第3范式?

- 学生(学号，系号，系主任)
  - **候选键 ：学号 $\xrightarrow{f}$ U；非主属性：系主任**
  - **传递依赖： 学号→系号，系号→系主任**
  - **无部分依赖**

  所以：满足第2NF但不满足第3NF.
- 员工(员工码，姓名，部门，部门经理)
  - **候选键 ：员工码 $\xrightarrow{f}$ U；非主属性：部门经理**
  - **传递依赖： 员工码→部门，部门→部门经理**
  - **无部分依赖**

  所以：满足第2NF但不满足第3NF.

➢ **关系模式设计如满足第3范式，则一定能满足第2范式；反之则不然。**

课后思考：举出一些满足第3范式和不满足第3范式但满足第2范式的实例

关系模式分解成3NF

示例：**R(A, B, C, D, E, F, G)**

函数依赖集合{ A→B, A→C, C→D, C→E, E→FG }

候选键：A; 有传递依赖，R不满足3NF。

分解规则：

将每一个函数依赖单独组成一个关系

ρ={ R1(A, B), R2(A, C), R3(C, D) , R4(C, E) , R5(E, F, G) }

可以看出：每一个模式都属于3NF

也可以合并一些关系：

ρ={ R12(A, B, C), R34(C, D, E) , R5(E, F, G) }

[Definition] BCNF

若R(U，F)$\in$1NF, 若对于任何$X \rightarrow Y \in F$ (或$X \rightarrow A \in F$), 当$Y \not\subset X$ (或$A \notin X$)时，X必含有候选键，则称R(U)属于Boyce-Codd范式，记为：R(U)$\in$BCNF。

示例：邮编(城市, 街道, 邮政编码)

函数依赖：{ 城市,街道 } → 邮政编码; 邮政编码→城市.

候选键 ：{ 城市,街道 }$\xrightarrow{f}$ U

因不含候选键：邮政编码→城市；所以不满足BCNF

因无传递依赖，所以满足第3范式；

示例: 选课(学号, 课程号, 教师编号)

假设规定每位教师只开一门课, 则有: { 学号,课程号 }→教师编号; 教师编号→课程号. 显然：该模式满足第3范式但不满足Boyce-Codd范式。

[定理]若R(U,F) $\in$BCNF, 则R(U,F)$\in$3NF。

证明：用反证法证明，设R(U,F) $\in$BCNF，但R(U,F)$\notin$3NF，依据3NF定义，则必有一传递依赖存在：

设该传递依赖为X$\rightarrow$Y，Y$\rightarrow$A, 其中X候选键，A$\notin$X, A$\notin$Y, Y$\not\rightarrow$X, 显然X$\not\subset$Y,

因A$\notin$Y, 则Y$\rightarrow$A将违反BCNF的定义(任一函数依赖都包含候选键，而Y不是候选键)。故定理得证。证毕。

➢**有传递依赖的或者说不满足3NF的，也一定不满足BCNF**

课后思考：举出一些满足BCNF和不满足BCNF但满足第3范式的实例

关系模式分解成BCNF

示例：R(A, B, C, D, E, F, G)

函数依赖集合{ A→B, A→C, C→D, C→E, E→FG }

候选键：A; 有不依赖于候选键的其他函数依赖，R不满足BCNF。

分解规则：

将左侧不含候选键的函数依赖单独组成一个关系, 将包含候选键的组成一关系

ρ={ R1(C, D), R2(C, E), R3(E, F, G) , R4(A, B, C) }

可以看出：R1 ∈BCNF; R2 ∈BCNF; R3 ∈BCNF; R4 ∈BCNF;

也可以将R1和R2合并：

ρ={ R12(C, D, E), R3(E, F, G) , R4(A, B, C) }

多值依赖

战德臣
哈尔滨工业大学　教授.博士生导师
黑龙江省教学名师
教育部大学计算机课程教学指导委员会委员

**Research Center on Intelligent
Computing for Enterprises & Services,
Harbin Institute of Technology**

[Definition]多值依赖

对R(U)，设X，Y⊆U，若对于R(U)的任一关系r，若元组t∈r，s∈r，t[X] = s[X]，则必有u∈r，v∈r使得：

(1) u[X]=v[X]=t[X]=s[X]

(2) u[Y]=t[Y]且u[U-X-Y] = s[U-X-Y]

(3) v[Y]=s[Y]且v[U-X-Y] = t[U-X-Y]

均成立，则称Y多值依赖于X，或说X多值决定Y，记作X→→Y。

	X	Y	Z=U-X-Y	
1	t[X]	t[Y]	t[Z]	t
2	t[X]	s[Y]	s[Z]	s
3	t[X]	t[Y]	s[Z]	u
4	t[X]	s[Y]	t[Z]	v

战德臣 教授

多值依赖的特性

1)直观地，对于X给定值，Y有一组值与之对应(0或n个)且这组Y值不以任何方式与U-X-Y中属性值相联系，有X→→Y。

2)若交换t, s 的Y值而得到的新元组仍在r中，则X→→Y。

3)X, Y不必不相交，u,v可以与t,s相同。

4)函数依赖是多值依赖的特例。

5)令Z=U-X-Y,有X→→Z, 若Z=ϕ, 则必有X→→Y。

示例：R = { 课程名C, 教师名T, 上课时间H, 教室R, 学生名S, 成绩G}，则有：

- C →→HR, T→→HR， 但不存在 C →→H及C →→R。

说明：同一门课程或同一教师对同一批学生可以在不同时间不同地点上课。

关于多值依赖的公理

战德臣
哈尔滨工业大学　教授.博士生导师
黑龙江省教学名师
教育部大学计算机课程教学指导委员会委员

Research Center on Intelligent
Computing for Enterprises & Services,
Harbin Institute of Technology

[Armstrong's Axioms A4~A8]关于多值依赖的公理

设R(U), X, Y⊆U, 对于R(U)的任一关系r, 有以下规则：

- [A4]多值依赖互补律(Complementation)或对称性：若X→→Y, 则X →→ U–X–Y；
- [A5] 多值依赖增广律(Augmentation)：若X→→Y且V⊆W, 则WX→→VY；

注意：此条与A2规则是相似的： X→Y且V⊆W，则WX→VY；

- [A6]多值依赖传递律(Transtivity)：若X→→Y, Y→→Z, 则X→→Z–Y；

注意：此条比A3规则限制要强：X→Y，Y→Z，则X→Z。多值依赖不存在这种规则，即：X→→Y，Y→→Z，则X→→Z不一定成立，例如C→→HR, HR→→H但是C不能多值决定H。

- [A7]若X→Y, 则X→→Y；
- [A8]若X→→Y，Z⊆Y且对于某个与Y不相交的W有W→Z, W∩Y=ϕ, 则有X→Z。

战德臣 教授

[定理]Armstrong Axioms系统的规则A1-A8是有效的

A6的证明：用反证法进行。

设一关系r, 假设X→→Y, Y→→Z, 而X→→Z−Y不成立，按多值依赖定义，即对任一关系r, 有元组t∈r, s∈r，但满足下述条件的u不存在(u∉r)：

u[X]=t[X]=s[X]，u[Z-Y]=t[Z-Y]且u[U-X-(Z-Y)]=s[U-X-(Z-Y)]

下面由X→→Y, Y→→Z检验上述u是否真的不存在：

由X→→Y, 对t∈r, s∈r，有v∈r, 满足：

(1) v[X]=t[X]=s[X]

(2) v[Y]=s[Y]

(3) v[U-X-Y]=t[U-X-Y]

由Y→→Z, 对v∈r, s∈r，有w∈r, 满足：

(4) w[Y]=v[Y]=s[Y]

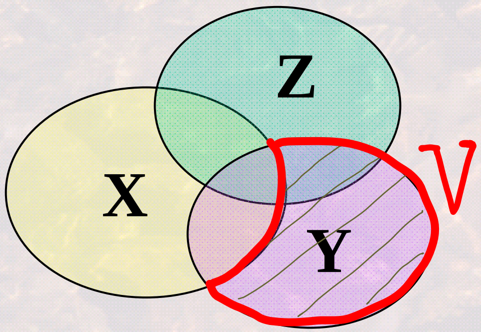
(5) w[Z]=v[Z]

(6) w[U-Z-Y]=s[U-Z-Y]

战德臣 教授

[定理]Armstrong Axioms系统的规则A1-A8是有效的

由X→→Y, 对t∈r, s∈r，有v∈r, 满足:
(1) v[X]=t[X]=s[X]
(2) v[Y]=s[Y]
(3) v[U-X-Y]=t[U-X-Y]
由Y→→Z, 对v∈r, s∈r，有w∈r, 满足:
(4) w[Y]=v[Y]=s[Y]
(5) w[Z]=v[Z]
(6) w[U-Z-Y]=s[U-Z-Y]



A6的证明：用反证法进行(Cont.)

由(1)~(6)可确定如下结论：

w[X]=t[X] (因为 X={Z∩X}∪{X-Z}, 对Z∩X，可由(5)(1)得到结果；对X-Z，可由(4)(6)(1)得到结果)

w[Z-Y]=t[Z-Y] (由(5)(1)(3)可得到结果)

w[U-X-(Z-Y)]=s[U-X-(Z-Y)] (设V=U-X-(Z-Y), V={V∩Z}∪{V-Z}, 对V-Z，可由(4)(6)得到结果；对V∩Z，由于V∩Z =(Y∩Z)-X，可由(2)(5)得到结果)

所以w就是u, 也就是说u是存在的，这与假设u不存在相矛盾，所以A6规则是正确的。

战德臣 教授

[定理]Armstrong Axioms系统的规则A1-A8是有效的

[A8]若X→→Y，Z⊆Y且对于某个与Y不相交的W有W→Z, W∩Y=ϕ, 则有X→Z。

A8的证明：用反证法。设一关系r满足X→→Y, W→Z, Z⊆Y, W∩Y=ϕ, 但X→Z不成立，按函数依赖定义，即对关系r, 有元组t∈r, s∈r，满足t[X]=s[X]但t[Z]≠s[Z]。

下面检验上述当t[X]=s[X]时是否有t[Z]≠s[Z]：

由X→→Y, 对t∈r, s∈r，有u∈r, 满足：

(1) u[X]=t[X]=s[X]

(2) u[Y]=t[Y]

(3) u[U-X-Y]=s[U-X-Y]

因为W∩Y=ϕ，由(1)(3)知u[W]=s[W],

又因W→Z，所以可推出 s[Z] = u[Z]，

又因 Z⊆Y, 由(2)可知u[Z]=t[Z]

所以有s[Z]= t[Z]，与假设s[Z] ≠t[Z]相矛盾，所以A8规则是正确的。

[引理7]：由Armstrong's Axioms可推出如下结论。

- (a)多值依赖合并律(Union Rule)：若X→→Y且X→→Z, 则X →→YZ。
- (b)多值依赖伪传递律(Pseudo Transitivity)：若X→→Y且WY→→Z,则X→→Z–WY。
- (c)混合伪传递律：若X→→Y, XY→Z, 则X→Z–Y
- (d)多值依赖分解律(Decomposition Rule)：若X→→Y，X→→Z则X→→Y–Z, X→→Z–Y, X→→Y∩Z。

➢证明：(略)

关系的第4范式和弱第4范式

战德臣

哈尔滨工业大学　教授.博士生导师

黑龙江省教学名师

教育部大学计算机课程教学指导委员会委员

Research Center on Intelligent Computing for Enterprises & Services, Harbin Institute of Technology

[Definition]4NF

设R(U)∈1NF，D是其上的一组依赖(函数依赖，多值依赖)，对任意X→→Y∈D，若Y≠ ϕ，Y⊄X，XY≠U，必有X为超键，则称R(U)满足第四范式，记为：R(U)∈4NF。

第四范式消除了非主属性对候选键以外属性的多值依赖。

如果有多值依赖，则一定依赖于候选键

[定理]若R∈4NF, 则必有R∈BCNF。

证明：设R∈4NF, 对R上的任何X→Y，Y−X ≠ φ,

(1)当XY = U时，X→U，X必为超键。

(2)当XY ≠ U时，因X→Y，有X→→Y，由第四范式定义X必为超键，再由BCNF定义知R∈BCNF。

[定理]若R上仅存在函数依赖，则若有R∈BCNF 即有R∈4NF, 反之，若R∈4NF，也有 R∈BCNF。

战德臣 教授

[Definition]W4NF

设R(U)∈3NF，若R上的任何互补多值依赖X→→Y(XY≠U，Y−X≠ϕ)和X→→(R−X−Y)中必有一个是函数依赖，则称R是弱第四范式的，记为R∈W4NF。

注：W4NF不一定是BCNF, 反之亦然。

回顾本讲学了什么?

战德臣

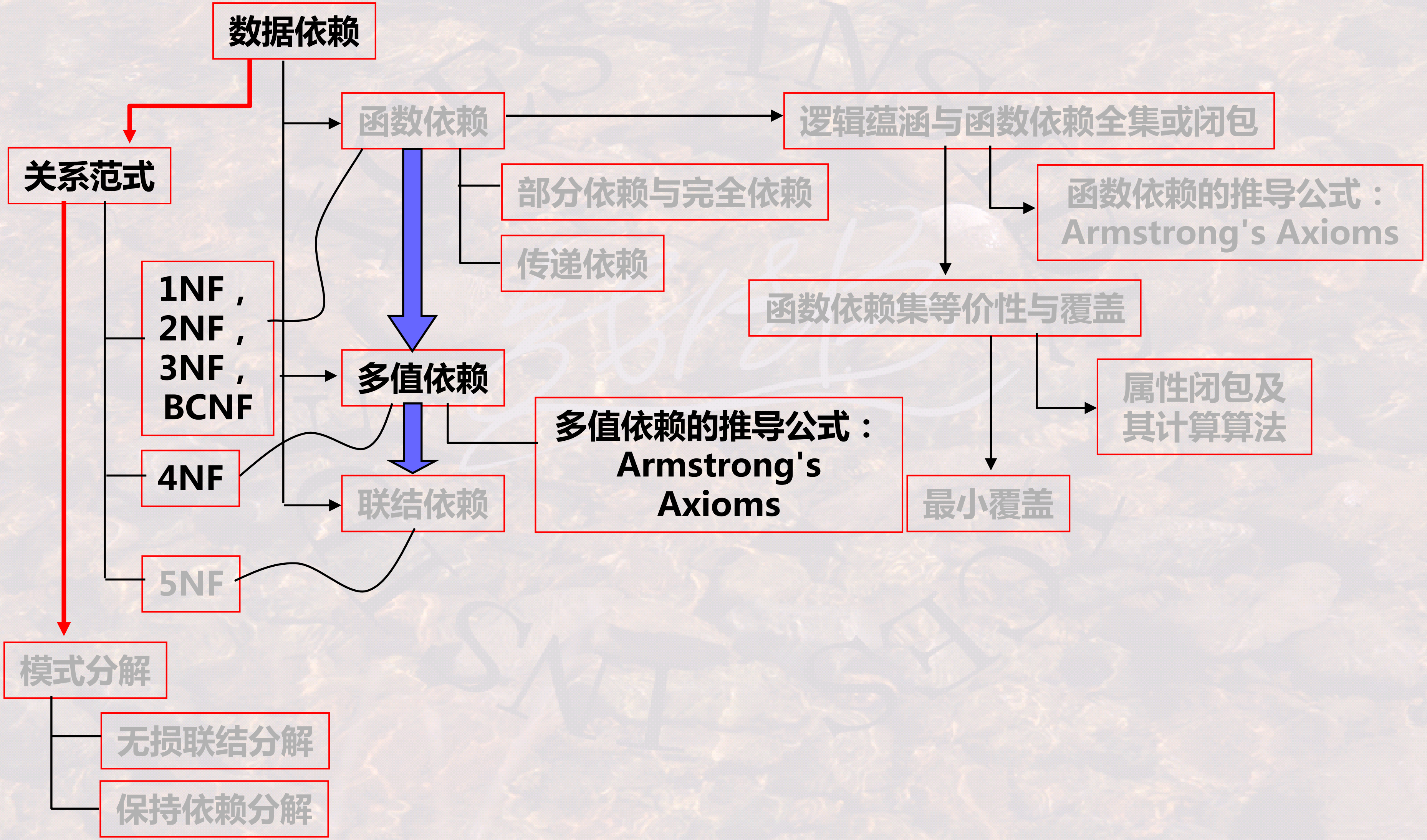
哈尔滨工业大学　教授.博士生导师

黑龙江省教学名师

教育部大学计算机课程教学指导委员会委员

Research Center on Intelligent Computing for Enterprises & Services, Harbin Institute of Technology

回顾本讲学习了什么?



数据依赖
关系范式
函数依赖
逻辑蕴涵与函数依赖全集或闭包
部分依赖与完全依赖
函数依赖的推导公式：
Armstrong's Axioms
传递依赖
1NF，
2NF，
3NF，
BCNF
函数依赖集等价性与覆盖
多值依赖
属性闭包及
其计算算法
多值依赖的推导公式：
Armstrong's
Axioms
4NF
联结依赖
最小覆盖
5NF
模式分解
无损联结分解
保持依赖分解