

# 第 3 章数据库系统

## 3.2 数据库模式与范式

### 3.2.1 数据库的结构与模式

1. 【2009 年题 4 解析】

本题考查对数据完整性约束方面基础知识的掌握。完整性约束为实体完整性约束、参照完整性约束和用户自定义完整性约束三类。其中实体完整性约束可以通过 Primary Key 指定，参照完整性约束通过 Foreign Key 指定，某些简单的约束可以通过 Check、Assertion 等实现。针对复杂的约束，系统提供了触发器机制，通过用户编程实现。本题中的约束条件只能通过编写职工表上的触发器，在对工资进行修改或插入新记录时触发，将新工资值与工资范围表中职工职务对应的工资范围对比，只有在范围内才提交，否则回滚。

【答案】C。

### 3.2.2 关系代数

1. 【2010 年题 7 解析】

【答案】B。

2. 【2011 年题 4 解析】

解答本题需要对关系代数中的自然连接有一定了解。自然连接操作会自动以两个关系模式中共有属性值相等作为连接条件，对于连接结果，将自动去除重复的属性。所以在本题中，连接条件为两个表的学号相等，当连接操作完成以后，形成的结果表，有属性“学号，姓名，年龄，入学时间，联系方式，课程号，成绩”，此时要选择 1 号课程的学生记录，应使用条件  $6 = "1"$ ，其含义是表中的第 6 个属性值为 “1” 。所以本题应选 B。

3. 【2013 年题 4 解析】

【答案】：C、B。

4. 【2014 年题 3 解析】

A 选项对应的是传递律

B 选项对应的是自反律

C 选项对应的是合并规则

D 选项对应的是分解规则(C 从 B 中分解出来，构成函数依赖)。

5. 【2014 年题 4 解析】

自然连接后，(A, R.B, R.C, D, E, F)，6 列。

$\sigma_{3 < 6}$ ，自然连接后，选取第 3 列分量 < 第 6 列分量所在行。

【答案】B、A、B。

6. 【2015 年题 5 解析】

【第一问解析】

A	B	C	D
6	3	1	5
6	5	7	4
6	3	7	4

【第二问解析】：(解析、计算结果)

其中  $R \times S$  结果为

A	B	C	D	C	D
6	3	1	5	1	5
6	3	1	5	7	4
6	1	5	1	1	5
6	1	5	1	7	4
6	5	7	4	1	5
6	5	7	4	7	4
6	3	7	4	1	5
6	3	7	4	7	4

$\sigma_{C=5}(R \times S)$ : 上述结果中行的第 3 个属性等于第 6 个属性，即  $R.C=S.D$ ，结果为：

A	B	C	D	C	D
6	1	5	1	1	5

$\pi_{A,D}(R)$ : 对上述结果投影第 1 个和第 4 个属性列，即 R 中的 A (记作 A)、R 中的 D (记作 R.D)，结果为

A	D
6	1

## 7. 【2016 年题 6 解析】

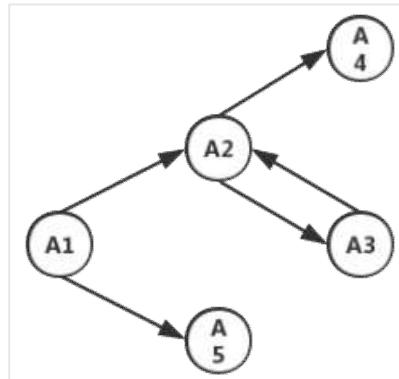
R 与 S 进行自然连接后，结果属性集为：A,B,C,D,E,F,G。投影操作后，结果为：A,B,D,F。

## 8. 【2016 年题 7 解析】

第一问：

解法一：A1 只出现在左边，是候选关键字；A4、A5 只出现在右边，不是候选关键字。且 A1 的闭包等于 R。所以 A1 为候选关键字。

解法二：通过绘制函数依赖图可以了解到，从 A1 出发，可以遍历全图，所以候选关键字为 A1。



第二问：

函数依赖 ( )  $\in F^+$ ：通俗一点，就是从 F 函数依赖集能推导出来的依赖关系。

A. A5  $\rightarrow$  A1A2    B. A4  $\rightarrow$  A1A2    C. A3  $\rightarrow$  A2A4    D. A2  $\rightarrow$  A1A5

根据函数依赖图可以看出 C 选项能走通。

## 9. 【2018 年题 4 解析】

注意本题中 R 与 S 是做自然联接操作，操作时会将 R 与 S 中相同字段名做等值联接，并将结果集去重复。所以 R 与 S 自然联接后的结果包括以下属性：R.A, R.B, R.C, D, E,

F, G。关系代数选择条件为“1”关系代数投影操作条件为“1,2,4,6,7”，对应的属性为：R.A, R.B, D, F, G。答案 BC。

#### 10. 【2018 年题 5 解析】

答案：B。

### 3.2.4 数据的规范化

#### 1. 【2009 年题 5 解析】

##### 【解析】

###### 第一问：

C 只出现在左边，是候选候选键；只有 C 选项包含 C，经验证 CE 能推导出 U。

因为  $E \rightarrow A \rightarrow B \rightarrow D$ ，则  $DE \rightarrow B$  也可推导出， $CE \rightarrow CE$ 。则 ABCDE 都可被推导出，即 CE 的闭包为 U。

###### 第二问：以此题为例

(1)：判断分解 p 是否为无损连接：若关系模式 R(U,F) 中，被分解为  $p=\{R_1, R_2\}$  是 R 的一个分解，若  $R_1 \cap R_2 \rightarrow R_1 - R_2$  或者  $R_1 \cap R_2 \rightarrow R_2 - R_1$ ，则为无损连接，此方法只适用于分解后的关系模式只有两个。

(2)：当关系模式是多个时候。

###### A 选项：

第一步：构造一个初始的二维表，模式中含有属性值的，记为  $\alpha_i$ ，i 为所在列数；不含有属性值的，记为  $b_{ij}$ ，其中 i 为所在行数，j 为所在列数。

属性 模式	A	B	C	D	E
R1(AC)	a1	b13	a3	b14	b15
R2(E)	b21	b22	b23	b24	a5
R3(DB)	b31	a2	b33	a4	b35

###### 第二步：

根据  $F=\{A \rightarrow B, DE \rightarrow B, CB \rightarrow E, E \rightarrow A, B \rightarrow D\}$  依次进行标识法判断。例如： $A \rightarrow B$

判断首先标识出 AB 所在列，发现  $A \rightarrow B$  中的决定因素 A 没有两行是相同的。再继续判断  $DE \rightarrow B, CB \rightarrow E, E \rightarrow A, B \rightarrow D$ 。

由于  $A \rightarrow B, DE \rightarrow B, CB \rightarrow E, E \rightarrow A, B \rightarrow D$  的决定因素中没有两行是相同的，因此选项 A 是有损连接的。B 选项类似。

###### C 选项：

属性 模式	A	B	C	D	E
R1(AC)	a1	b13	a3	b14	b15
R2(E)	b21	b22	b23	a4	a5
R3(DB)	a1	a2	b33	b34	b35

发现  $A \rightarrow B$  中的决定因素 A 的第 1 行与第 3 行的值相同，将列 B 第 1 行变成 a2（这里的判断依据是：列 B 第 1 行与第 3 行中如果有  $\alpha_i$ ，则 B 第 1 行与第 3 行都变成  $\alpha_i$ ；如果没有，则取行号最小值，假如列 B 第 3 行为 b32，则 B 第 1 行与第 3 行都变成 b13(行号最小)）。通过规则转换如下：

属性 模式	A	B	C	D	E

R1(AC)	a1	a2	a3	b14	b15
R2(E)	b21	b22	b23	a4	a5
R3(DB)	a1	a2	b33	b34	b35

再变换  $DE \rightarrow B$  决定因素中没有两行是相同的；继续判断  $CB \rightarrow E$ ，发现决定因素中没有两行是相同的；再判断  $E \rightarrow A$ ，发现  $E \rightarrow A$  决定因素中没有两行是相同的；继续判断  $B \rightarrow D$ ，发现列 B 的第 1 行与第 3 行的值相同。则将 D 的第 3 行变成 b14(依据为：没有  $\alpha_i$  就取行号最小的值)。转换为

属性 模式	A	B	C	D	E
R1(AC)	a1	a2	a3	b14	b15
R2(E)	b21	b22	b23	a4	a5
R3(DB)	a1	a2	b33	b14	b35

### 第三步：

反复检查函数依赖集 F，无法修改上表，发现上表中没有一行是  $a1, a2, a3, a4, a5$ 。则是有损连接。

### D 选项：

属性 模式	A	B	C	D	E
R1(ABC)	a1	a2	a3	b14	b15
R2(ED)	b21	b22	b23	a4	a5
R3(ACE)	a1	b32	a3	b34	a5

发现  $A \rightarrow B$  中的决定因素 A 的第 1 行与第 3 行的值相同，将列 B 第 3 行变成 a2(依据为：没有  $\alpha_i$  就取行号最小的值)。

属性 模式	A	B	C	D	E
R1(ABC)	a1	a2	a3	b14	b15
R2(ED)	b21	b22	b23	a4	a5
R3(ACE)	a1	a2	a3	b34	a5

再变换  $DE \rightarrow B$ ，发现决定因素中没有两行是相同的；继续判断  $CB \rightarrow E$ ，发现  $CB \rightarrow E$  中的决定因素 CB 的第 1 行与第 3 行的值相同，则将 A 的第 1 行变成 a5(依据为：没有  $\alpha_i$  就取行号最小的值)。转换成如下：

属性 模式	A	B	C	D	E
R1(ABC)	a1	a2	a3	b14	a5
R2(ED)	b21	b22	b23	a4	a5
R3(ACE)	a1	a2	a3	b34	a5

继续判断  $E \rightarrow A$ ，发现  $E \rightarrow A$  中的决定因素 E 的第 2 行与第 3 行的值相同，则将 A 的第 2 行变成 a1(依据为：没有  $\alpha_i$  就取行号最小的值)。转换成如下：

属性 模式	A	B	C	D	E
R1(ABC)	a1	a2	a3	b14	b15
R2(ED)	a1	b22	b23	a4	a5
R3(ACE)	a1	a2	a3	b34	a5

再判断  $B \rightarrow D$ ，发现发现  $B \rightarrow D$  中的决定因素 B 的第 1 行与第 3 行的值相同，则将列 D

第 3 行变成 b14(依据为：没有  $\alpha_i$  就取行号最小的值)。

属性 模式	A	B	C	D	E
R1(ABC)	a1	a2	a3	b14	b15
R2(ED)	a1	b22	b23	a4	a5
R3(ACE)	a1	a2	a3	b14	a5

发现上表中没有一行 a1,a2,a3,a4,a5。

反复检查函数依赖集  $F=\{A \rightarrow B, DE \rightarrow B, CB \rightarrow E, E \rightarrow A, B \rightarrow D\}$ ，看根据已知是否能推导出其他关系。发现由  $\{A \rightarrow B, CB \rightarrow E\}$  可推导出  $AC \rightarrow E$ ；  $\{E \rightarrow A, A \rightarrow B, B \rightarrow D\}$  可推导出  $E \rightarrow D$ 。

首先看  $AC \rightarrow E$  中的决定因素 AC 的第 1 行与第 3 行的值相同(同时为 a1,a3)，则将列 E 第 1 行变成 a5(依据为：没有  $\alpha_i$  就取行号最小的值)。

属性 模式	A	B	C	D	E
R1(ABC)	a1	a2	a3	b14	a5
R2(ED)	a1	b22	b23	a4	a5
R3(ACE)	a1	a2	a3	b14	a5

再看  $E \rightarrow D$  中的决定因素 E 的第 1、2、3 行的值相同，则将列 D 第 1、2、3 行变成 a4(依据为：没有  $\alpha_i$  就取行号最小的值)。

属性 模式	A	B	C	D	E
R1(ABC)	a1	a2	a3	a4	a5
R2(ED)	a1	b22	b23	a4	a5
R3(ACE)	a1	a2	a3	a4	a5

发现上表中第 1 行为 a1,a2,a3,a4,a5。判断为无损连接，其实第 3 行 a1,a2,a3,a4,a5，只要有一行满足条件即可。

I:保持函数依赖的分解保持函数依赖，就是指原来有哪些函数依赖，当进行拆分以后，这些函数依赖在新的关系模式中，是否依然存在。

如原来有关系模式 P(C, S, T, R, G)，函数依赖为：  $F=\{C \rightarrow T, ST \rightarrow R, TR \rightarrow C, SC \rightarrow G\}$ 。分解成  $P1(C, T, R), P2(C, S, G)$ ，

其中保持了  $P1(C, T, R)$ :  $C \rightarrow T, TR \rightarrow C$  函数依赖。

$P2(C, S, G)$  保持了：  $SC \rightarrow G$  函数依赖。

结果函数依赖：  $ST \rightarrow R$  就丢了。所以没有保持。

此题函数依赖有问题，不具体解析。

【答案】C、D。

## 2. 【2009 年题 6 解析】

关系模型的完整性规则是对关系的某种约束条件。关系模型中可以有三类完整性约束：实体完整性、参照完整性和用户定义的完整性。实体完整性规定基本关系的主属性不能取空值。

由于①仓库关系 W 中的“负责人”引用员工关系的员工号，所以应满足参照完整性约束；②库存关系 I 中的“仓库号，产品号”唯一标识 I 中的每一个记录，所以应满足实体

完整性约束；③ 职称为“工程师”的月薪不能低于 3500 元，是针对某一具体关系数据库的约束条件，它反映某一具体应用所涉及的数据必须满足的语义要求，所以应满足用户定义完整性约束。因此，试题(1)的正确答案为 B。

SQL 查询是数据库中非常重要的内容。该 SQL 查询要求对查询结果进行分组，即具有相同名称的产品的元组为一组，然后计算每组的库存数量。由此可排除 A、B 和 D，所以试题(2)正确答案为 C。

### 3. 【2011 年题 3 解析】

题目虽然有多个问题，但实际上只考查了一个知识点——实体之间的联系。

供应关系中，有属性：项目号，零件号，供应商号。这些属于分别来自供应商、项目、零件这三个关系，并且，一个供应商可以向多个项目供应零件，一个供应商可以供应多种零件，一个项目可以由多个供应商供应零件，一个项目可以使用多种零件，而一种零件可以由多个不同供应商来提供，一种零件可用于不同项目。这说明供应关系涉及 3 个实体，这 3 个实体之间的关系是 k:n:m。

从题目的描述“若一个工程项目可以有多个员工参加，每个员工可以参加多个项目”可以得知，项目和员工的关系是 n:m。

在实体转关系模式过程中，存在 3 种类型的联系，他们的处理方式如下：

1:1 联系：在两个关系模式中的任意一个模式中，加入另一个模式的键和联系类型的属性；

1:n 联系：在 n 端实体类型对应的关系模式中加入 1 端实体类型的键和联系类型的属性；

m:n 联系：将联系类型也转换成关系模式，属性为两端实体类型的键加上联系类型的属性。

试题中是 m:n 联系，所以需要把联系单独转成一个关系模式。

**【答案】**D、C、A。

### 4. 【2012 年题 41 解析】

本题考查的是应试者关系数据库方面的基础知识。

根据题意，零件 P 关系中的(商品代码，供应商)可决定的零件 P 关系的所有属性，所以零件 P 关系的主键为(商品代码，供应商)；又因为，根据题意(商品代码，供应商)→商品名称，而商品代码→商品名称，供应商→联系方式，可以得出商品名称和联系方式都部分依赖于码(存在非主属性对码的部分函数依赖)，所以，该关系模式属于 1NF。

关系模式 P 属于 1NF，1NF 存在冗余度大、修改操作的不一致性、插入异常和删除异常四个问题。所以需要对模式分解，其中选项 A、选项 B 和选项 C 的分解是有损且不保持函数依赖。例如，选项 A 中的分解 P1 的函数依赖集 F1=Φ，分解 P2 的函数依赖集 F2=Φ，丢失了 F 中的函数依赖，即不保持函数依赖。

**【答案】**A、D。

### 5. 【2013 年题 3 解析】

R 有函数依赖集 F={A→B, B→C}。由于 A 可确定 B 和 C，所以 A 为主键，单个属性的主键不可能有部分依赖关系，所以 R 已符合 2NF。进一步分析是否为 3NF 时，需要识别 R 中是否存在传递依赖。A→B, B→C 属于典型的传递依赖，所以 R 最高只到 2NF。

当 R 被拆分为 R1 与 R2 后，R1 与 R2 分别只有两个属性，此时的关系模式不可能存在部分依赖，也没法传递依赖（至少 3 个属性才可能传递），所以都达到了 3NF。

接下来判断是否无损分解，由于：U1 ∩ U2=A, U1-U2=B, U2-U1=C。

而 R 中有函数依赖：A→B，所以分解是无损分解。

最后判断是否保持函数依赖：

R1 中包含 A 与 B 两个属性，所以 A→B 依赖关系被 R1 保持下来了。

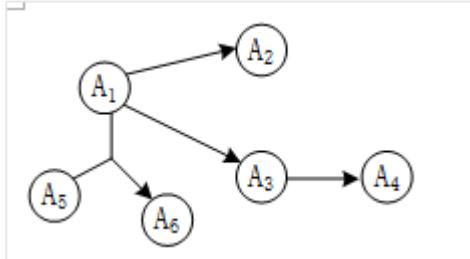
而 R2 中的 A 与 C 两个属性，没有保持任何函数依赖，导致函数依赖 B→C 丢失，所以分解没有保持函数依赖。

**【答案】**：D、D。

## 6. 【2017 年题 7 解析】

要求关系模式的候选码，可以先将函数依赖画成图的形式：

从图很直观的可以看出，入度为零的结点是 A1 与 A5，从这两个结点的组合出发，能遍历全图，所以 A1A5 组合键为候选码。



题目后一问是一个概念性问题，2NF 的规定是消除非主属性对码的部分函数依赖。本题已明确告知未消除该依赖，说明未达到 2NF，只能选 1NF。

【答案】：C、A。

## 3.3 数据库设计

### 1. 【2012 年题 42 解析】

数据库设计主要分为用户需求分析、概念结构、逻辑结构和物理结构设计四个阶段。其中，在用户需求分析阶段中，数据库设计人员采用一定的辅助工具对应用对象的功能、性能、限制等要求所进行的科学分析，并形成需求说明文档、数据字典和数据流程图。用户需求分析阶段形成的相关文档用以作为概念结构设计的设计依据。

【答案】C、B。

### 3.3.3 需求分析

#### 1. 【2009 年题 3 解析】

需求分析阶段的任务是对现实世界要处理的对象（组织、部门和企业等）进行详细调查，在了解现行系统的概况，确定新系统功能的过程中收集支持系统目标的基础数据及处理方法。需求分析是在用户调查的基础上，通过分析，逐步明确用户对系统的需求。在需求分析阶段应完成的文档是数据字典和数据流图。

【答案】C。

### 3.3.5 逻辑结构设计

#### 1. 【2010 年题 5 解析】

数据库设计分为用户需求分析、概念设计、逻辑设计和物理设计四个主要阶段。将抽象的概念模型转化为与选用的 DBMS 产品所支持的数据模型相符合的逻辑模型，它是物理设计的基础。包括模式初始设计、子模式设计、应用程序设计、模式评价以及模式求精。

逻辑设计阶段的任务是将概念模型设计阶段得到的基本 E-R 图，转换为与选用的 DBMS 产品所支持的数据模型相符合的逻辑结构。如采用基于 E-R 模型的数据库设计方法，该阶段就是将所设计的 E-R 模型转换为某个 DBMS 所支持的数据模型；如采用用户视图法，则应进行模式的规范化，列出所有的关键字以及用数据结构图描述表集合中的约束与联系，汇总各用户视图的设计结果，将所有的用户视图合成一个复杂的数据库系统。

【答案】C。

## 3.4 事务管理

### 3.4.1 并发控制

#### 1. 【2012 年题 14 解析】

本题主要考查数据库系统架构设计知识。在数据库系统中，“事务”是访问并可能更新各种数据项的一个程序执行单元。为了保证数据完整性，要求数据库系统维护事务的原子性、一致性、隔离性和持久性。

题干中第 1 个架构设计场景描述了数据库设计中为了实现原子性和持久性的最为简单的策略：“影子拷贝”。该策略假设在某一个时刻只有一个活动的事务，首先对数据库做副本(称为影子副本)，并在磁盘上维护一个 dp\_pointer 指针，指向数据库的当前副本。对于要执行写操作的数据项，数据库系统在磁盘上维护数据库的一个副本，所有的写操作都在数据库副本上执行，而保持原始数据库不变，如果在任一时刻操作不得不中止，系统仅需要删除新副本，原数据库副本没有受到任何影响。

题干中的第 2 个架构设计场景主要考查考生对事务一致性实现机制的理解。事务的一致性要求在没有其它事务并发执行的情况下，事务的执行应该保证数据库的一致性。数据库系统通常采用完整性约束检查机制保证单个事务的一致性。

题干中的第 3 个架构设计场景主要考查数据库的锁协议。两阶段锁协议是实现事务隔离性的常见方案，该协议通过定义锁的增长和收缩两个阶段约束事务的加锁和解锁过程，能够保证事务的串行化执行，但由于事务不能一次得到所有需要的锁，因此该协议可能会导致死锁。

题干中的第 4 个架构设计场景主要考查数据库的恢复机制，主要描述了基于日志的延迟修改技术(deferred-modification technique)的设计与恢复过程。该技术通过在日志中记录所有对数据库的修改操作，将一个事务的所有写操作延迟到事务提交后才执行，日志中需要记录“事务开始”和“事务提交”时间，还需要记录数据项被事务修改后的值，无需记录数据项被事务修改前的原始值。当系统发生故障时，如果某个事务已经开始，但没有提交，则该事务对数据项的修改尚未体现在数据库中，因此无需做任何恢复动作。

【答案】B、C、A、D、C。

#### 1. 【2015 年题 3 解析】

题目所描述的情况为环路等待，此时系统处于死锁状态。

【答案】A。

## 3.6 分布式数据库系统

### 3.6.1 分布式数据库的概念

#### 1. 【2015 年题 4 解析】

本题考查对分布式数据库基本概念的理解。

分片透明是指用户或应用程序不需要知道逻辑上访问的表具体是怎么分块存储的。

复制透明是指采用复制技术的分布方法，用户不需要知道数据是复制到哪些节点，如何复制的位置透明是指用户无须知道数据存放的物理位置。

逻辑透明，即局部数据模型透明，是指用户或应用程序无须知道局部场地使用的是哪种数据模型。

【答案】D。

#### 2. 【2017 年题 9 解析】

所谓的两个阶段是指：第一阶段：准备阶段(表决阶段)和第二阶段：提交阶段(执行阶

段)。

准备阶段(表决阶段)：事务协调者(事务管理器)给每个参与者(资源管理器)发送 Prepare 消息，每个参与者要么直接返回失败(如权限验证失败)，要么在本地执行事务，写本地的 redo 和 undo 日志，但不提交，到达一种“万事俱备，只欠东风”的状态。

提交阶段(执行阶段)：如果协调者收到了参与者的失败消息或者超时，直接给每个参与者发送回滚(Rollback)消息；否则，发送提交(Commit)消息；参与者根据协调者的指令执行提交或者回滚操作，释放所有事务处理过程中使用的锁资源。(注意：必须在最后阶段释放锁资源)

**【答案】：**C。

## 3.7 数据仓库

### 1. 【2018 年题 6 解析】

数据仓库 4 大特点：

- ① 面向主题：数据按主题组织。
- ② 集成的：消除了源数据中的不一致性，提供整个企业的一致性全局信息。
- ③ 相对稳定的(非易失的)：主要进行查询操作，只有少量的修改和删除操作（或是不删除）。
- ④ 反映历史变化(随着时间变化)：记录了企业从过去某一时刻到当前各个阶段的信息，可对发展历程和未来趋势做定量分析和预测。

答案 C。

## 3.8 数据挖掘

### 1. 【2013 年题 16 解析】

数据挖掘的任务有关联分析、聚类分析、分类分析、异常分析、特异群组分析和演变分析，等等。

## 3.11 其他

### 1. 【2016 年题 8 解析】

从题目中“系统中有正在运行的事务”可知应采用动态方式，从题目中“全部数据”可知应是全局转储，所以应采用：动态全局转储。

### 2. 【2017 年题 8 解析】

题目中表达式：存在从关系 R 中选择的元组 t 的 C 列上的分量，大于关系 S 中的一个元组 u 在 B 列上的分量。

$t[3] < u[2]$ : R 中每行的第三个分量 (R 的第 3 列) < S 中每行的第二个分量

$t[3]=\{3, 6, 9, 12\}$ ,  $u[2]=\{7, 5, 9, 10\}$

$t[3]$  中的  $3 < \{7, 5, 9, 10\}$  中的  $7, 5, 9, 10$ ，满足要求。

$t[3]$  中的  $6 < \{7, 5, 9, 10\}$  中的  $7, 9, 10$ ，满足要求。

$t[3]$  中的  $9 < \{7, 5, 9, 10\}$  中的  $10$ ，满足要求。

$t[3]$  中的  $12$  不满足要求。存在：只要满足  $u[2]$  中一个分量就行。

所以  $t[3] < u[2] = \{(1, 2, 3), (4, 5, 6), (7, 8, 9)\}$

**【答案】：**C。