

## 6.1 问题的提出

## 6.2 规范化 (重点)

## 6.3 数据依赖的公理系统 (自学)

## 6.4 模式的分解 (了解)

## 6.1 问题的提出

---

### 建立关系数据库的**首要任务**

对于一个具体问题，如何构造适合于它的**数据库模式**？

- 建立数据库应构造几个关系？
- 每个关系有什么属性？

### 什么是一个**好的数据库逻辑设计**

## 6.1 问题的提出

□ 一个例子，开发一个学校教务的数据库，涉及的对象有：

学号(*Sno*),系(*Sdept*),系主任姓名(*Mname*),课号(*Cno*),成绩(*Grade*)

### ■ 语义：

- 1、一个系有若干学生，但一个学生只属于一个系；
- 2、一个系只有一名主任；
- 3、一个学生可选修多门课程，每门课程有若干学生选修；
- 4、每个学生所学的每门课程都有一个成绩。

建立数据库应构造几个表？每个表有什么属性？

### ■ 为存储并管理这些信息，设计一个关系模式：

*Student(Sno, Sdept, Mname, Cno, Grade)*

## 关系模式 *Student* 主码 (*sno, cno*)

<i>Sno</i>	<i>Sdept</i>	<i>Mname</i>	<i>Cno</i>	<i>Grade</i>
20001001	IS	张三	101	82
20001001	IS	张三	102	85
20001221	IS	张三	101	90
20003001	MS	王五	104	80
20003001	MS	王五	106	78

❑ 关系模式 *Student* 存在的问题： **不是一个好的关系模式**

■ **数据冗余大、更新复杂、插入异常、删除异常、**

- **更新复杂**：若更换系主任，须修改与该系学生有关的每个元组。
- **插入异常**：若新成立系，还没有招生，就无法把这个系及其系主任的信息存入数据库。
- **删除异常**：若某系的学生全部毕业了删除该系学生信息的同时，把这个系及其系主任的信息也丢掉了。

## 6.1 问题的提出

---

### □ 什么是一个好的模式？

- 不会发生插入、删除和更新异常、数据冗余少。

### □ 问题的原因

- 模式中的某些**数据依赖**引起的。

### □ 什么是**数据依赖**？

- 一个关系内部属性与属性之间的约束关系
- 反映现实世界属性间的相互联系

## 6.1 问题的提出

---

□ 本章中把**关系模式**看作是一个**三元组**： $R\langle U, F \rangle$ 。

■  $R$ ：关系名， $U$ ：属性的集合， $F$ ：数据依赖

□ 上例关系模式可记为 $Student\langle U, F \rangle$ ：

■  $U = \{ Sno, Sdept, Mname, Cno, Grade \}$

■  $F = \{ Sno \rightarrow Sdept, Sdept \rightarrow Mname, (Sno, Cno) \rightarrow Grade \}$

□ 如何解决关系模式中存在的问题？

■ **规范化理论**：消除关系模式中**不合适的数据依赖**，  
以解决插入异常、删除异常、更新异常和数据冗余问题

## 6.2.1 函数依赖

### 1. 函数依赖

设 $R(U)$ 是属性集 $U$ 上的关系模式。 $X, Y$ 是 $U$ 的子集。若对于 $R(U)$ 的任意一个可能的关系 $r$ ,  $r$ 中不可能存在两个元组在 $X$ 上的属性值相等,而在 $Y$ 上的属性值不等,则称 $X$ 函数确定 $Y$ 或 $Y$ 函数依赖于 $X$ , 记作 $X \rightarrow Y$ 。 $X$ 称为这个函数依赖的决定属性组。

### 2. 平凡函数依赖与非平凡函数依赖

决定因素

□ 如果 $X \rightarrow Y$ , 而且 $Y \not\subseteq X$ , 则称为非平凡的函数依赖。

□ 如果 $Y \subseteq X$ , 则称 $X \rightarrow Y$ 为平凡的函数依赖。

**非平凡函数依赖:**  $(Sno, Cno) \rightarrow Grade$

**平凡函数依赖:**  $(Sno, Cno) \rightarrow Sno, (Sno, Cno) \rightarrow Cno$

对任一关系模式,  
平凡依赖必然存在

### 3. 完全函数依赖与部分函数依赖

□ 关系模式 $R(U)$ 中, 如果 $X \rightarrow Y$ , 而且对于 $X$ 的任何一个真子集 $X'$ , 都有 $X' \not\rightarrow Y$ , 则称 $Y$ 完全函数依赖于 $X$ , 记作 $X \xrightarrow{f} Y$ 。

□ 若 $X \rightarrow Y$ , 但是 $Y$ 不完全函数依赖于 $X$ , 则称 $Y$ 部分函数依赖于 $X$ , 记作 $X \xrightarrow{p} Y$ 。

■  $(Sno, Cno) \xrightarrow{f} Grade$ , 属于**完全**函数依赖。

■  $(\textcolor{red}{Sno}, Cno) \xrightarrow{p} Sdept$ , 属于**部分**函数依赖。

因为  $Sno \rightarrow Sdept$



### 4. 传递函数依赖

□ 在 $R(U)$ 中, 如果 $X \rightarrow Y$  ( $Y \not\subseteq X$ ),  $Y \not\rightarrow X$ ,  $Y \rightarrow Z$ , 且 $Z \not\subseteq Y$ , 则称

$Z$  传递函数依赖于 $X$ 。记作 $X \xrightarrow{\text{传递}} Z$ 。

■  $Sno \rightarrow Sdept, Sdept \rightarrow Mname, Sno \xrightarrow{\text{传递}} Mname$

### 5. 码

□ 设 $K$ 为关系模式 $R\langle U, F \rangle$ 中的属性或属性组合。若

$K \xrightarrow{f} U$ , 则 $K$ 称为 $R$ 的一个**候选码**。若关系模式 $R$ 有多个候选码, 则选定其中的一个作为**主码**。

## 6.2.3 范式

---

- **范式**是符合某一种级别的关系模式的集合。
- 关系数据库中的关系必须满足一定的要求。满足不同程度要求的为不同范式。
- 范式的种类和联系
  - $1NF \supset 2NF \supset 3NF \supset BCNF \supset 4NF \supset 5NF$
  - 某一关系模式 $R$ 为第 $n$ 范式，可简记为 $R \in nNF$

**规范化：**一个低一级范式的关系模式，通过模式分解可以转换为若干个高一级范式的关系模式的集合。

## 6.2.3 范式

□ **第一范式**：如果一个关系模式 $R$ 的所有属性都是**不可分**的基本数据项，则 $R \in 1NF$ 。

- 第一范式是对关系模式的**最低要求**。不满足第一范式的数据库模式不能称为关系数据库。

满足 $1NF$ 的关系模式是不是一个好的关系模式。?

不一定

例：关系模式  $Student = (\underline{Sno}, \underline{Cno}, Grade, Sdept, Mname)$

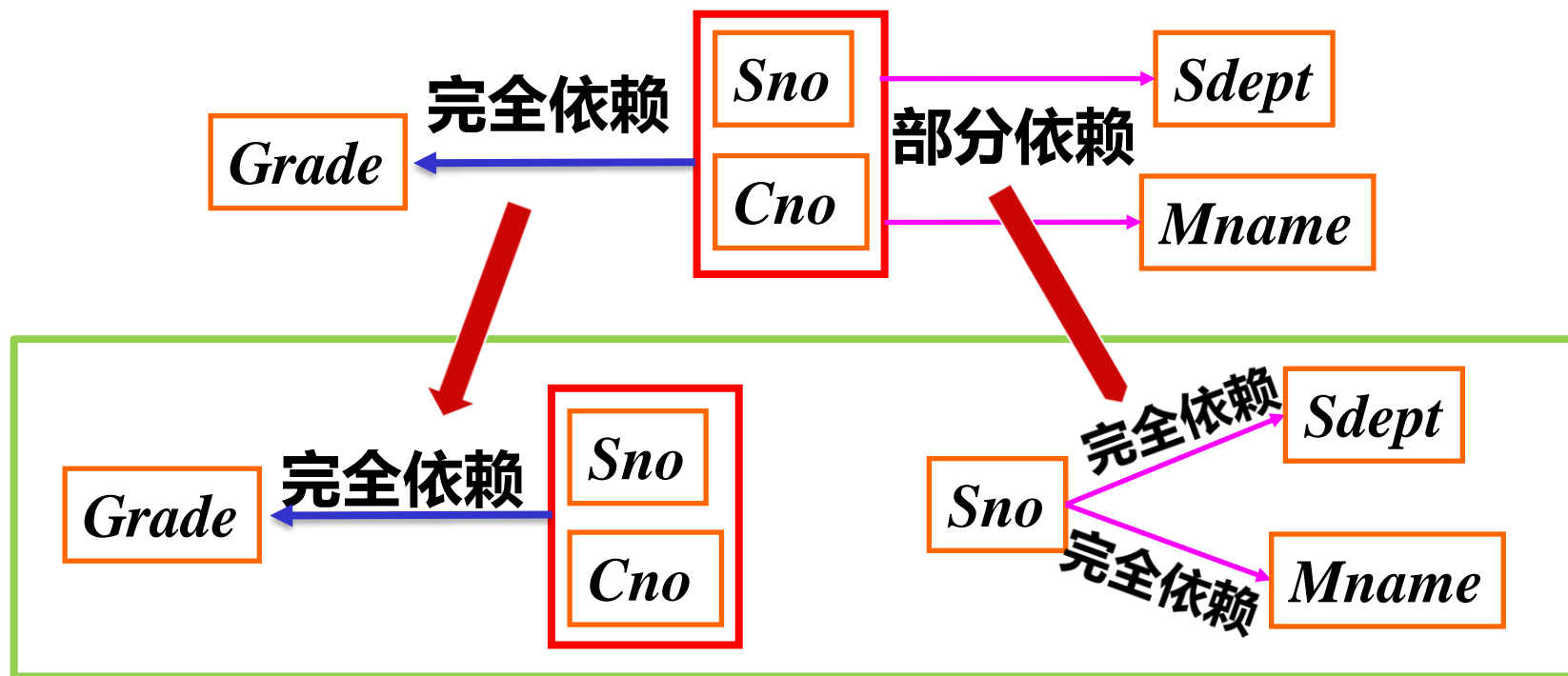
- 是第一范式。 **但不是一个好的关系模式。**
- 码为 $(Sno, Cno)$ 。
- **非主属性** $Sdept$ 和 $Mname$ **部分函数依赖于码** $(Sno, Cno)$

## 6.2.4 2NF

□ 若关系模式 $R \in 1NF$ ，且**每一个非主属性都完全函数依赖**于 $R$ 的码，则 $R \in 2NF$ 。

关系模式 $Student = (\underline{Sno}, \underline{Cno}, Grade, Sdept, Mname)$ **不属于2NF**

解决方法：采用投影分解法分解 $Student$ ，消除部分函数依赖



$SC(\underline{Sno}, \underline{Cno}, Grade)$

$SM(\underline{Sno}, Sdept, Mname)$

**将非第二范式的关系模式分解为若干个第二范式关系模式的方法：**

- 1、把关系模式中对码完全函数依赖的非主属性与决定它们的码放在一个关系模式中。**
- 2、把对码部分函数依赖的非主属性和决定它们的主属性放在一个关系模式中。**
- 3、检查分解后的新模式，若仍然不是2NF，则继续按前面的方法进行分解，直到达到要求为止。**

## 6.2.4 2NF

2NF是否解决了1NF存在的问题 

关系模式  $SC(\underline{Sno}, Cno, Grade) \in 2NF$ ,  $SM(\underline{Sno}, Sdept, Mname) \in 2NF$

<i>Sno</i>	<i>Cno</i>	<i>Grade</i>
20001001	101	82
20001002	102	85
20001003	101	90

<i>Sno</i>	<i>Sdept</i>	<i>Mname</i>
20001001	IS	张三
20001002	IS	张三
20001003	IS	张三

□ 2NF 未能完全消除关系模式中的各种异常情况和数据冗余

■ 原因：存在传递依赖

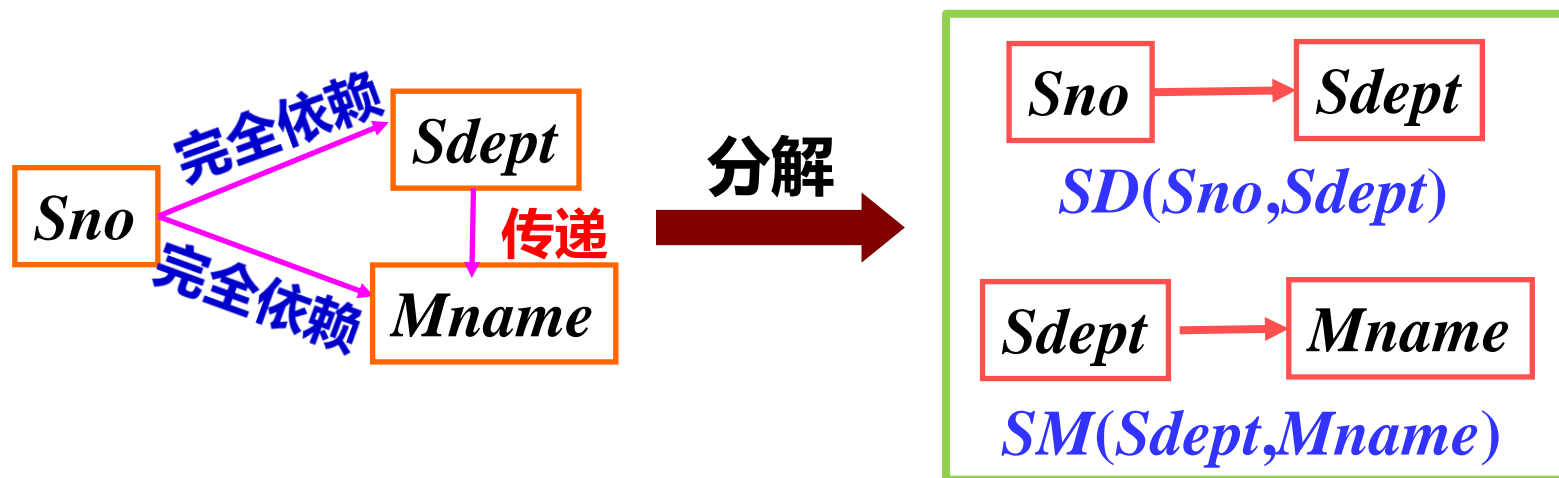
○  $Sno \rightarrow Sdept, Sdept \rightarrow Mname, Sno \xrightarrow{\text{传递}} Mname$

## 6.2.5 3NF

□ 若 $R \in 2NF$ ，且每个非主属性都不传递依赖于码，则 $R \in 3NF$ 。

关系模式 $SC(\underline{Sno}, Cno, Grade) \in 3NF$ ,  $SM(\underline{Sno}, Sdept, Mname) \notin 3NF$


解决方法：采用投影分解法分解 $SM$ ，消除传递函数依赖



□ 至此，关系模式 $Student$ 已被分解为如下三个关系模式：  
 $SC(Sno, Cno, Grade)$ 、 $SD(Sno, Sdept)$ 、 $SM(Sdept, Mname)$   
均属于第三范式。

将非第三范式的关系模式分解为若干个第三范式关系模式的方法：

- 1、把关系模式中**直接对码函数依赖的非主属性**与决定它们的**码**放在一个关系模式中。
- 2、把造成**传递函数依赖的决定因素**连同被它们**决定的属性**放在一个关系模式中。
- 3、检查分解后的新模式，若仍然不是3NF，则继续按前面的方法进行分解，直到达到要求为止。

3NF是否彻底解决了2NF存在的问题？



# 一个例子

□ 关系模式  $STC(S, T, C)$ ,  $S$ : 学生;  $T$ : 教师;  $C$ : 所教课程

■ 假设每个教师只教一门课程,  $T \rightarrow C$

■ 每门课程由多个老师教

■ 某一学生选定某门课就确定了教师,  $(S, C) \rightarrow T$

■ 某一学生选修某个教师的课就确定了课程,  $(S, T) \rightarrow C$

$STC \in \textcircled{?} NF$

$S$	$T$	$C$
1	王涛	英语
1	李明	体育
1	张霞	高数
2	王涛	英语
3	李明	体育
3	王涛	英语

$STC$ 的候选码  $\textcircled{?}$

□  $(S, C)$ 和 $(S, T)$ 都是候选码

□  $STC$ 没有非主属性, 属于 $3NF$

有无异常  $\textcircled{?}$

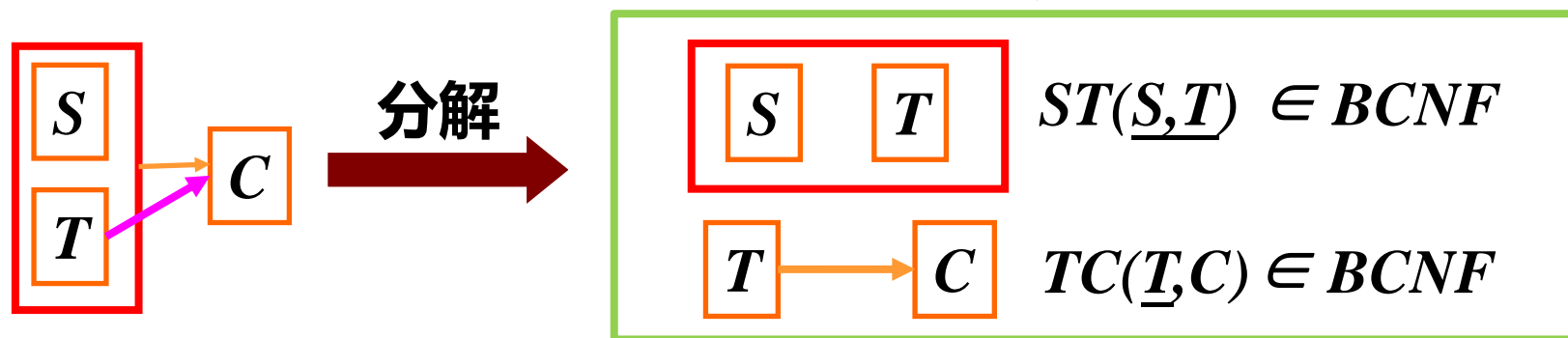
仍有异常

原因: 主属性 $C$ 部分依赖于码 $(S, T)$

## 6.2.6 BC范式 (BCNF)

- 对关系模式  $R \in 1NF$ ，若  $X \rightarrow Y$  且  $Y \not\subseteq X$  时  $X$  必含有码，则  $R \in BCNF$ 。 **每一个决定因素都包含码**

■  $STC$  中，  $T \rightarrow C$ ，  $T$  未包含码，  $STC \notin BCNF$



- 如果  $R \in 3NF$ ，且  $R$  只有一个候选码，则  $R$  必属于  $BCNF$ 。
- 如果一个关系数据库中所有关系模式都属于  $BCNF$ ，那么在函数依赖范围内，它已实现了模式的彻底分解。
- $BCNF$  是基于函数依赖的最高范式，但不是数据库模式设计的最高范式。

## 6.2.6 BC范式 (BCNF)

---

将3NF的关系模式分解为BCNF关系模式的方法：

- 1、在各个3NF模型中，去掉一些属性，只保留主码，使它们只有**唯一的候选码**。
- 2、把从各个模型中去掉的**主属性**，分别同各自的**决定属性**组成新的关系模式。
- 3、检查分解后的新模式，若仍然不是BCNF，则继续按前面的方法进行分解，直到达到要求为止。

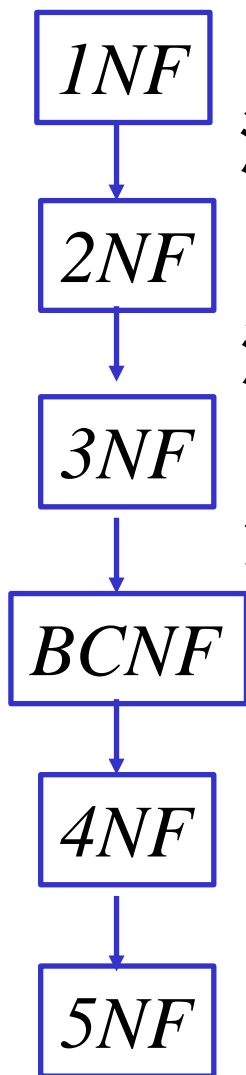
## 6.2.7 关系模式的规范化

- 关系数据库的规范化理论是数据库逻辑设计的工具。
- 属性都是不可分的数据项的关系就是最基本的规范化关系。
- 关系模式的规范化过程：一个低一级范式的关系模式通过模式分解可以转换为若干高一级范式的关系模式集合。

模式分解的原则：分解后的关系模式与原关系模式等价。

- 判断等价的标准有三种：
  - 分解具有无损连接性。若将关系模式 $A$ 分解为关系模式 $B$ 和 $C$ ，则要求 $B$ 与 $C$ 进行自然连接的结果等于 $A$ 。
  - 分解要保持函数依赖。
  - 分解既要保持函数依赖，又要具有无损连接性。

## 6.2.7 关系模式的规范化



消除**非主属性**对码的**部分**函数依赖。

消除**非主属性**对码的**传递**函数依赖。

消除**主属性**对码的**部分和传递**函数依赖。

□ **规范化的基本思想：**逐步消除数据依赖中不合适的部分，使各关系模式达到某种程度的“分离”。

□ **规范化的实质：**概念的单一化。

□ **不是规范化程度越高的关系模式就越好，上面的规范化步骤可以在其中任何一步终止。**

## 6.2.7 关系模式的规范化

---

- 1、确定所有的**候选码**；
- 2、确定**主码**；
- 3、确定**主属性**和**非主属性**。
- 4、找出属性间的依赖关系（**函数依赖**）；
- 5、根据应用特点，确定规范到**第几范式**；
- 6、分解关系，力求分解后的各关系相互独立。

1、有关系模式  $A(C, T, H, R, S)$ ，函数依赖集：

$F = \{C \rightarrow T, (H, R) \rightarrow C, (H, T) \rightarrow R, (H, S) \rightarrow R\}$ ， $A$ 的码是（**D**）。

A、 $C$               B、 $(H, R)$               C、 $(H, T)$               D、 $(H, S)$

2、关系模式 $R$ 的整个属性全是主属性，则 $R$ 满足的最高范式至少是（**D**）

A、 $2NF$               B、 $1NF$               C、 $BCNF$               D、 $3NF$

3、书P<sub>203</sub>页第7题

**7、下面的结论哪些是正确的，哪些是错误的？对于错误的结论请给出理由或给出一个反例说明之。**

**(1)任何一个二目关系都是属于3NF的。** ✓

**(2)任何一个二目关系都是属于BCNF的。** ✓

**(5)若 $R.A \rightarrow R.B$ ,  $R.B \rightarrow R.C$ , 则 $R.A \rightarrow R.C$**  ✓

**(6)若 $R.A \rightarrow R.B$ ,  $R.A \rightarrow R.C$ , 则 $R.A \rightarrow R.(B,C)$**  ✓

**(7)若 $R.B \rightarrow R.A$ ,  $R.C \rightarrow R.A$ , 则 $R.(B,C) \rightarrow R.A$**  ✓

**(8)若 $R.(B,C) \rightarrow R.A$ , 则 $R.B \rightarrow R.A$ ,  $R.C \rightarrow R.A$**  ✗



**4、设有一个反映工程及其所使用相关材料信息的关系模式：**

**R(工程号，工程名，工程地址，开工日期，完工日期，材料号，材料名称，使用数量)**

**规定每个工程的地址、开工、完工日期惟一；不同工程的地址、开工和完工日期可能相同；工程名与材料名称均有重名；每个工程使用若干种材料，每种材料可应用于若干工程中。**

**(1)根据上述规定，写出模式R的基本函数依赖和主码。**

**工程号->工程名，工程号->工程地址，**

**工程号->开工日期，工程号->完工日期，**

**材料号->材料名称，（工程号，材料号）->使用数量**

**主码为（工程号，材料号）**

**工程号->工程名, 工程号->工程地址,  
工程号->开工日期, 工程号->完工日期,  
材料号->材料名称, (工程号, 材料号) ->使用数量  
主码为 (工程号, 材料号)**

**(2)R最高达到第几范式, 并说明理由。**

**R最高达到1NF。**

**因为存在部分依赖, (工程号, 材料号) ->工程名**

**(3)将R规范到3NF。**

**R1 (工程号, 工程名, 工程地址, 开工日期, 完工日期)**

**R2 (材料号, 材料名称), R3 (工程号, 材料号, 使用数量)**