**E-R的设计原则**

* 尽量减小实体集，能作为属性时不要作为实体集。
* “属性”不能再具有需要描述的性质。“属性”必须是不可分割的数据项，不能包括其他属性。
* “属性”不能与其他实体具有联系。在E-R中所有的联系必须是实体间的联系，而不能有属性与实体之间的联系。
* 针对特定用户的应用，确定实体、属性和实体间的联系，设计该用户视图的局部E-R图。
* 综合局部E-R图，产生出总体E-R图。在综合过程中，同名实体只能出现一次，并去掉不必要的联系，以便消除冗余。一般来说，从总体E-R图必须能导出原来的所有局部视图，包括实体、属性和联系。

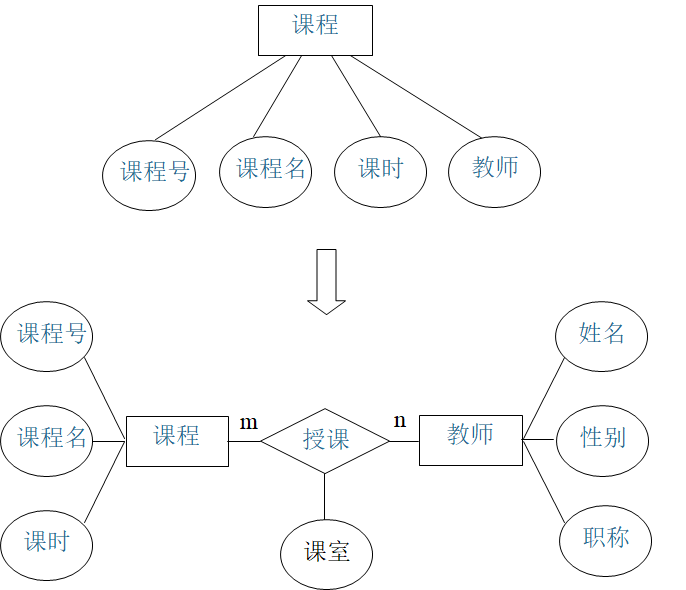
**E-R图设计的方法**

​ 在设计E-R图时，一般使用先局部后全局的方法。

* 选择局部应用：根据某个系统的具体情况，在多层的数据流图中选择一个适当层次的数据流图作为设计分E-R图的出发点 。
* 逐一设计分E-R图：将数据字典中的数据抽取出来，参照数据流图，设计出E-R图，再作必要的调整。
* 调整原则：为简化图的处置，现实世界中的事物能作为属性对待的，尽量作为属性对待。作为“属性”，不能再具有描述的性质，也不能与其他实体具有联系。

**E-R图的设计步骤**

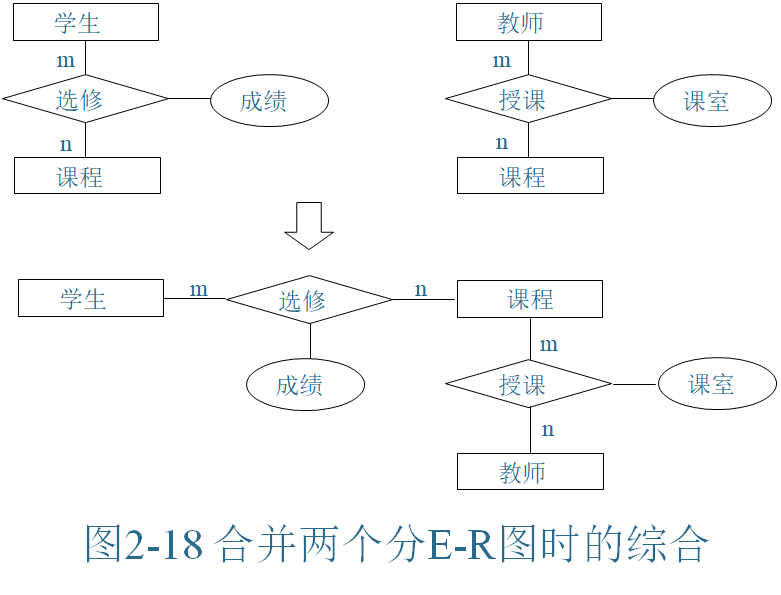
* 设计分E-R图
  + **首先选择局部应用。根据某个系统的具体情况，在多层的DFD中选择一个适当层次的DFD，作为设计分E-R图的出发点。**让这组图中每一部分对应一个局部应用。由于高层的DFD只能反映系统的概貌，而中层的DFD能较好地反映系统中各局部应用的子系统组成，因此**人们往往以中层DFD作为设计分E-R图的依据。**
  + 选择好局部应用之后，就要对每个局部应用逐一设计分E-R图。在前面选好的某一层次的DFD中，每个局部应用都对应了一组DFD，局部应用涉及的数据都已经收集在数据字典中了。现在就是要将这些数据从数据字典中抽取出来，参照DFD，标定局部应用中的实体、实体的属性、标识实体的主码，确定实体之间的联系及其类型。
  + 在实际E-R图设计时，实体与属性之间并不存在形式上可以截然划分的界限，现实世界的事物能作为属性对待的尽量作为属性对待，从自然划分的内容出发定义雏形的E-R图，再进行必要的调整。例如，图2-17所示的是一个有属性上升为用实体集表示的实例。



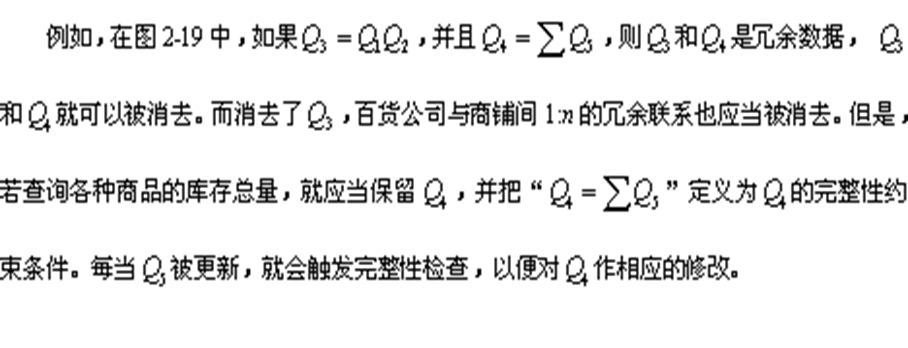
* 分E-R图的集成  
  各子系统的分E-R图设计好以后，下一步就是要将所有的分E-R图集成一个系统的总E-R图。一般说来，分E-R图的合成有两种方法：第一种方法是多个分E-R图一次集成；第二种方法是逐步集成，用累加的方式一次集成两个分E-R图。第一种方法比较复杂，做起来难度较大。第二种方法每次只集成两个分E-R图，可以降低复杂度。无论采用哪种方式，每次集成分E-R图时都需要分两步走：
  + **将各分E-R图合并起来生成初步E-R图**

**合并时注意解决各分E-R图之间的冲突。**由于各个局部应用所面向的问题是不同的，而且通常是由不同的设计人员进行不同局部的视图设计，这样就会导致各个分E-R图之间必定会存在许多不一致的地方，即产生冲突问题。由于各个分E-R图存在冲突，所以不能简单地把它们画到一起，必须先消除各个分E-R图之间的不一致，形成一个能被全系统所有用户共同理解和接受的统一的概念模型，再进行合并。合理消除各个分E-R图的冲突是进行合并的主要工作和关键所在。

* + - 分E-R图之间的冲突主要包括属性冲突、属性取值单位冲突和命名冲突三种。
    - 属性冲突 属性冲突主要有两种情况：
      * 属性域冲突，即属性值的类型、取值范围或取值集合不同。例如，对于零件号属性，不同的部门可能会采用不同的编码形式，而且定义的类型又各不相同，有的定义为整型，有的则定义为字符型，这都需要各个部门之间协商解决。
      * 属性取值单位冲突。例如，零件的重量，不同的部门可能分别用公斤、斤或千克来表示，结果会给数据统计造成错误。
    - 命名冲突 命名冲突有两种：
      * 同名异义冲突，即不同意义的对象在不同的局部应用中具有相同的名字。
      * 异名同义冲突，即意义相同的对象在不同的局部应用中有不同的名字。
    - 结构冲突 结构冲突有3种情况：
      * 同一对象在不同的应用中具有不同的抽象。例如，职工在某一局部应用中被当作实体对待，而在另一局部应用中被当作属性对待，这样就会产生抽象冲突问题。
      * 同一实体在不同分E-R图中的属性组成不一致，即所包含的属性个数和属性排列次序不完全相同。这类冲突是由于不同的局部应用所关心的实体的不同侧面而造成的。解决这类冲突的方法是使该实体的属性取各个分E-R图中属性的并集，再适当调整属性的次序，使之兼顾到各种应用。
      * 实体之间的联系在不同的分E-R图中呈现不同的类型。此类冲突解决方法是根据应用的语义对实体联系的类型进行综合或调整。



* 消除不必要的冗余，生成基本E-R图。
  + 在初步E-R图中可能存在冗余的数据和实体间冗余的联系。所谓冗余数据是指可由基本数据导出的数据。所谓冗余的联系是可由其他联系导出的联系。冗余的存在容易破坏数据库的完整性，给数据库维护增加困难，应当加以消除。消除了冗余的初步E-R图就称为基本E-R图。
  + 常见的消除冗余方法有分析方法和规范化理论方法。
  + 用分析方法消除冗余 分析方法是消除冗余的主要方法。分析方法消除冗余是以数据字典和DFD为依据，根据数据字典中关于数据项之间逻辑关系的说明来消除冗余的。
    - 在实际应用中，并不是要将所有的冗余数据与冗余联系都消除。有时为了提高数据查询效率、减少数据存取次数，在数据库中就设计了一些数据冗余或联系冗余。因而，在设计数据库结构时，冗余数据的消除或存在要根据用户的整体需要来确定。如果希望存在某些冗余，则应在数据字典的数据关联中进行说明，并把保持冗余数据的一致作为完整性约束条件。



* 用规范化理论消除冗余 基于第9章介绍的规范化理论消除冗余。

