

11 数据库设计

数据库设计概述

11.1.1 数据库应用系统的生命期

数据库应用系统的生命周期分为数据库规划、需求描述与分析、数据库与应用程序设计、数据库设计实现、测试、运行维护6个阶段。

1、数据库规划

数据库规划的任务是确定软件的开发目标及可行性。该阶段应该给出问题定义、可行性分析和项目开发计划。

2、需求描述与分析

需求描述与分析是以用户的角度，从系统中的数据和业务规则入手，收集和整理用户的信息，以特定的方式加以描述，是下一步工作的基础。

3、数据库与应用程序设计

数据库的设计是对用户数据的组织和存储设计；应用程序设计是在数据库设计基础上对数据操作及业务实现的设计，包括事务设计和用户界面设计。

4、数据库设计实现

数据库设计实现是依照设计，使用DBMS支持的数据定义语言实现数据库的建立，用高级语言编写应用程序。

数据库设计概述

5、测试

测试是在数据库系统投入使用之前，通过精心制订的测试计划和测试数据来测试系统的性能是否满足设计要求，以便发现问题。

6、运行维护

数据库应用系统经过测试、试运行后即可正式投入运行。运行维护是系统投入使用后，必须不断地对其进行评价、调整和修改，直至系统消亡。

11.1.2 数据库设计的一般策略

数据库设计的一般策略有两种：自顶向下和自底向上。

11.1.3 数据库设计的基本步骤

一般将数据库设计分为如下6个阶段：（1）用户需求分析。（2）概念结构设计。（3）逻辑结构设计。（4）物理结构设计。（5）数据库实施阶段。（6）数据库运行和维护阶段。

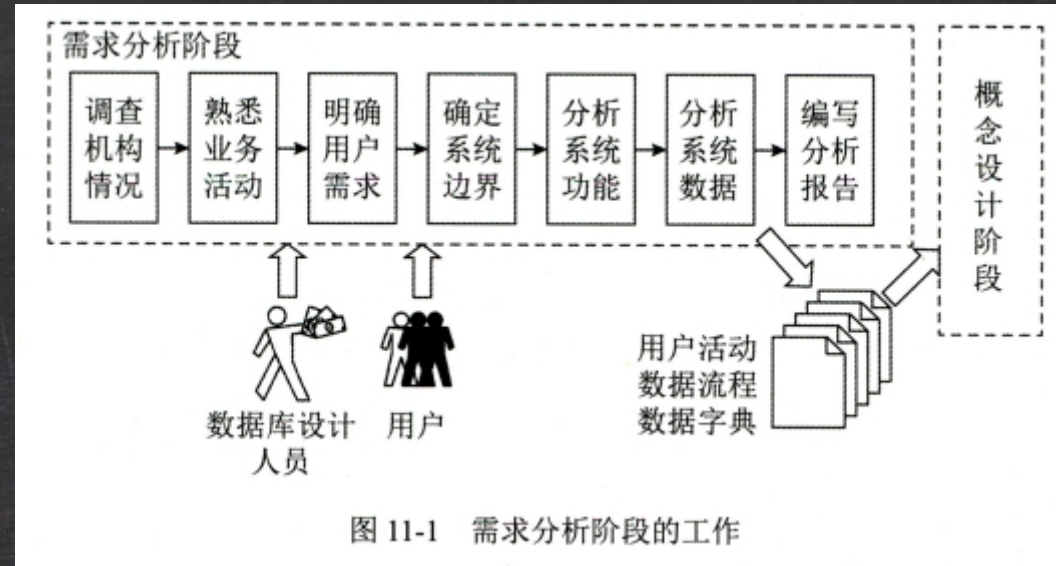
系统需求分析

11.2.1 需求分析的任务、方法和目标

需求分析阶段的任务：综合各个用户的应用需求，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。

- （1）**信息要求**。用户需要在系统中保存哪些信息，由这些保存的信息要得到什么样的信息，这些信息以及信息间应当满足的完整性要求。
- （2）**处理要求**。用户在系统中要实现什么样的操作功能，对保存信息的处理过程和方式，各种操作处理的频度、响应时间要求、处理方式等以及处理过程中的安全性要求和完整性要求。
- （3）**系统要求**。包括安全性要求、使用方式要求和可扩充性要求。安全性要求：系统有几种用户使用，每一种用户的使用权限如何。使用方式要求：用户的使用环境是什么，平均有多少用户同时使用，最高峰有多少用户同时使用，有无查询相应的时间要求等。可扩充性要求：对未来功能、性能和应用访问的可扩充性的要求。

系统需求分析



11.2.2 需求分析阶段的文档

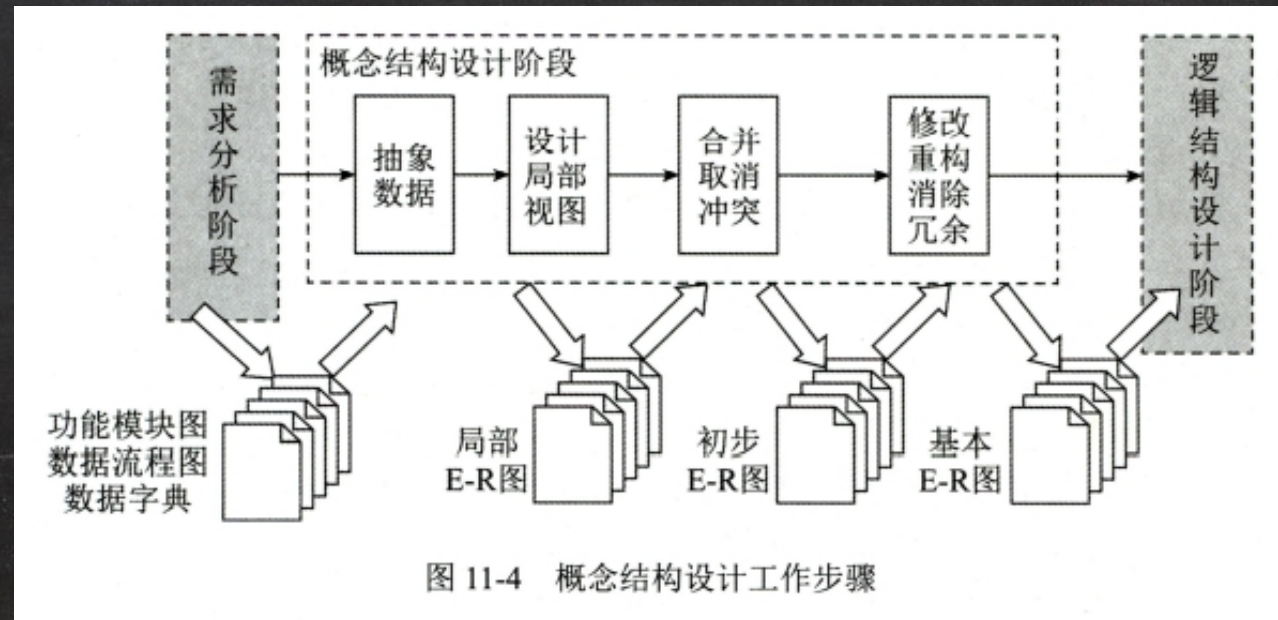
需求分析阶段的成果是系统需求说明书，主要包括数据流图、数据字典、各种说明性表格、统计输出表和系统功能结构图等。

数据字典 (Data Dictionary, DD) 是各类数据描述的集合，它是关于数据库中数据的描述，即元数据，而不是数据本身。如用户将向数据库中输入什么信息，从数据库中要得到什么信息，各类信息的内容和结构，信息之间的联系等。

数据字典包括数据项、数据结构、数据流、数据存储和处理过程5个部分。

概念结构设计

- 概念结构设计是在需求分析的基础上，依照需求分析中的信息要求，对用户信息加以分类、聚集和概括，建立信息模型。
- 概念结构设计最著名最常用的方法是实体-联系方法，简称E-R方法。它将现实世界的信息结构统一用实体、属性以及实体之间的联系来描述。
- 概念结构设计工作步骤包括：选择局部应用、逐一设计分E-R图和E-R图合并。



概念结构设计

1、选择局部应用

选择适当层次的数据流图，让这一层的每一部分对应一个局部应用，实现某一项功能，从这一层入手，就能很好地设计分E-R图。

2、逐一设计分E-R图

划分好各个局部应用之后，就要对每一个局部应用逐一设计分E-R图，又称为局部E-R图。

现实生活中许多事物，作为实体还是属性没有明确的界定，这需要根据具体情况而定，一般遵循以下两条准则：

- (1) 属性不可再分，即属性不再具有需要描述的性质，不能有属性的属性。
- (2) 属性不能与其他实体发生联系，联系是实体与实体间的联系。

3、E-R图合并

根据局部应用设计好各局部E-R图之后，就可以对各分E-R图进行合并。合并的目的在于解决冲突，消除冗余。

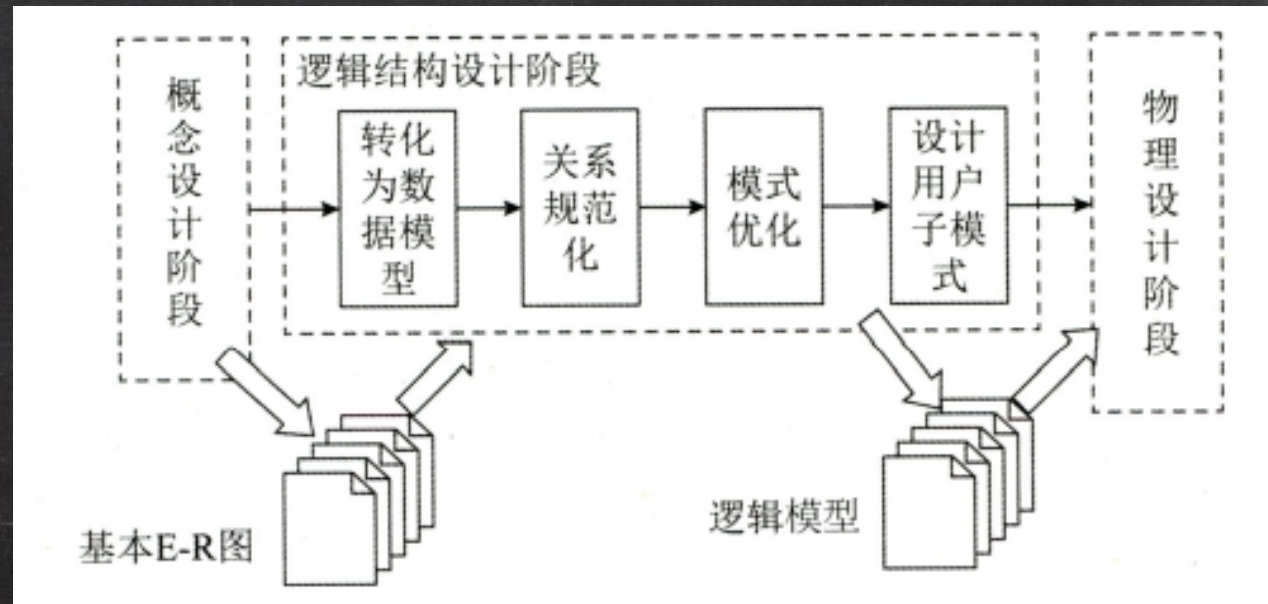
概念结构设计

分E-R图合并时，它们之间存在的冲突主要有以下三类：

- (1) **属性冲突**：同一属性可能会存在于不同的分E-R图中，由于设计人员不同或是出发点不同，对属性的类型、取值范围、数据单位等可能会不一致。
- (2) **命名冲突**：相同意义的属性，在不同的分E-R图上有着不同的命名，或是名称相同的属性在不同的分E-R图中代表着不同的意义，这些也需要进行统一。
- (3) **结构冲突**：同一实体在不同的分E-R图中有不同的属性，同一对象在某一分E-R图中被抽象为实体而在另一分E-R图中又被抽象为属性。

逻辑结构设计

- 逻辑结构设计就是在概念结构设计的基础上进行数据模型设计，可以是层次模型、网状模型和关系模型。
- 逻辑结构设计阶段的主要工作步骤包括确定数据模型、将E-R图转换成指定的数据模型、确定完整性约束和确定用户视图。



逻辑结构设计

11.4.1 E-R图向关系模式的转换

- 1、实体向关系模式的转换
 - 2、联系向关系模式的转换
- 重要，最后讲。

11.4.2 关系模式的规范化

由E-R图转换得来的初始关系模式并不能完全符合要求，还会有数据冗余、更新异常存在，这就需要经过进一步的规范化处理。

- (1) 根据语义确定各关系模式的数据依赖。
- (2) 根据数据依赖确定关系模式的范式，判断关系模式是否达到了3NF或4NF。
- (3) 如果关系模式不符合要求，要进行分解，达到3NF、BCNF或4NF。
- (4) 关系模式的评价及修正。有时根据处理要求，可能还需要增加部分冗余以满足处理要求，这就需要做部分关系模式的处理，分解、合并或增加冗余属性，提高存储效率和处理效率。

11.4.3 确定完整性约束

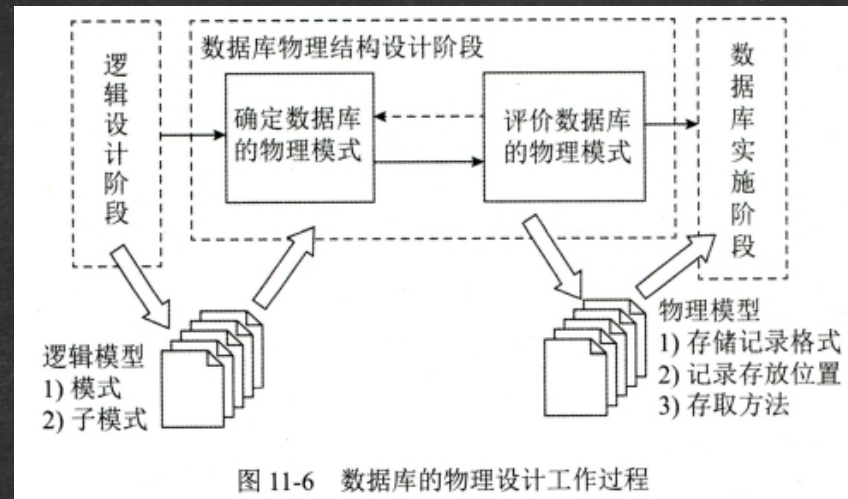
11.4.4 用户视图的确定

- (1) 根据数据流图确定处理过程使用的视图。
- (2) 根据用户类别确定不同用户使用的视图。

数据库的物理设计

数据库在物理设备上的存储结构与存取方法称为数据库的物理结构，它依赖于给定的计算机系统。为一个给定的逻辑数据模型设计一个最适合应用要求的物理结构的过程，就是数据库的物理设计。

11.5.1 数据库物理设计工作过程



11.5.2 数据库物理设计工作步骤

物理设计的主要工作步骤包括：确定数据分布、存储结构和访问方式。

数据库的物理设计

1、确定数据分布

- (1) 根据不同应用分布数据。
- (2) 根据处理要求确定数据的分布。
- (3) 对数据的分布存储必然会导致数据的逻辑结构的变化，要对关系模式做新的调整，回到数据库逻辑设计阶段做必要的修改。

2、确定数据的存储结构

存储结构具体指数据文件中记录之间的物理结构。

为提高数据的访问速度，通常会采用索引技术。在物理设计阶段，要根据数据处理和修改要求，确定数据库文件的索引字段和索引类型。

3、确定数据的访问方式

- (1) 存储记录结构设计。
- (2) 存储记录布局。
- (3) 存取方法的设计。

数据库运行维护与管理

1、数据库重组和重构

数据库重组是指在不改变数据库逻辑和物理结构的情况下，去除数据库存储文件中的废弃空间以及碎片空间中的指针链，使数据库记录在物理上紧连。

数据库系统运行过程中，会因为一些原因而对数据库的结构做修改，称为**数据库重构**。重构包括表结构的修改和视图的修改。

2、数据库系统的审计

审计是一种DBMS工具，它记录数据库资源和权限的使用情况。启用审计功能，可以产生审计跟踪信息，包括哪些数据库对象受到了影响，谁在什么时候执行了这些操作。

审计是被动的，它只能跟踪对数据库的修改而不能防止，但作为一个安全性手段，起到对非法入侵的威慑作用，可以据此追究非法入侵者的法律责任。

审计功能的开启会影响系统的性能。

数据库运行维护与管理

3、数据库的存储管理

在数据库系统运行过程中，随着数据的不断变更，会影响到系统的响应效率。通过以下手段进行存储管理，可有效地提高系统性能。

- (1) 索引文件和数据文件分开存储，事务日志文件存储在高速设备上。
- (2) 适时修改数据文件和索引文件的页面大小。
- (3) 定期对数据进行排序。
- (4) 增加必要的索引项。

4、数据安全性管理

- (1) 建立网络安全，主要是防火墙的设置。
- (2) 操作系统级安全，进行登录用户的管理。
- (3) DBMS级安全，对访问数据库的用户进行密码验证。
- (4) 角色和用户的授权管理。
- (5) 建立视图和存储过程加强安全性。
- (6) 使用审计功能，为追究非法入侵者法律责任提供证据，发现安全漏洞。

数据库运行维护与管理

5、SQL语句的编码检验

- (1) 尽可能地减少多表查询或建立物化视图。
- (2) 以不相关子查询替代相关子查询。
- (3) 只检索需要的列。
- (4) 用带IN的条件子句等价替换OR子句。
- (5) 经常提交COMMIT，以尽早释放锁。

6、数据安全性管理

- (1) 如果频繁的访问是对两个相关的表进行连接操作，则考虑将其合并。
- (2) 如果频繁的访问只是在表中的某一部分字段上进行，则考虑分解表，将该部分单独作为一个表。
- (3) 对于更新很少的表，引入物化视图。

7、索引维护和改进

- (1) 如果查询是瓶颈，则在关系上建立适应的索引，通常在作为查询条件的属性上建立索引，可以提高查询效率。
- (2) 如果更新是瓶颈，每次更新都会重建表上的索引，引起效率的降低，则考虑删除某些索引。
- (3) 选择适当的索引类型，如果是经常使用范围查询，则B树索引比散列索引更高效。
- (4) 将有利于大多数数据查询和更新的索引设为聚簇索引。

联系向关系模式的转换

1、一对一联系(1:1):

(1) 将联系转换成一个独立的关系模式，关系模式的名称取联系的名称，关系模式的属性包括该联系所关联的两个实体的码及联系的属性，关系的码取自任一方实体的码。

任职(班主任工号, 班级编号, 任职时间)

(2) 将联系归并到关联的两个实体的任意一方，给待归并的一方实体属性集中增加另一方实体的码和该联系的属性即可，归并后的实体码保持不变。

班主任(工号, 姓名, 身份证号, 住址, 班级编号, 任职时间)

或： 班级(班级编号, 名称, 人数, 班主任编号, 任职时间)



联系向关系模式的转换

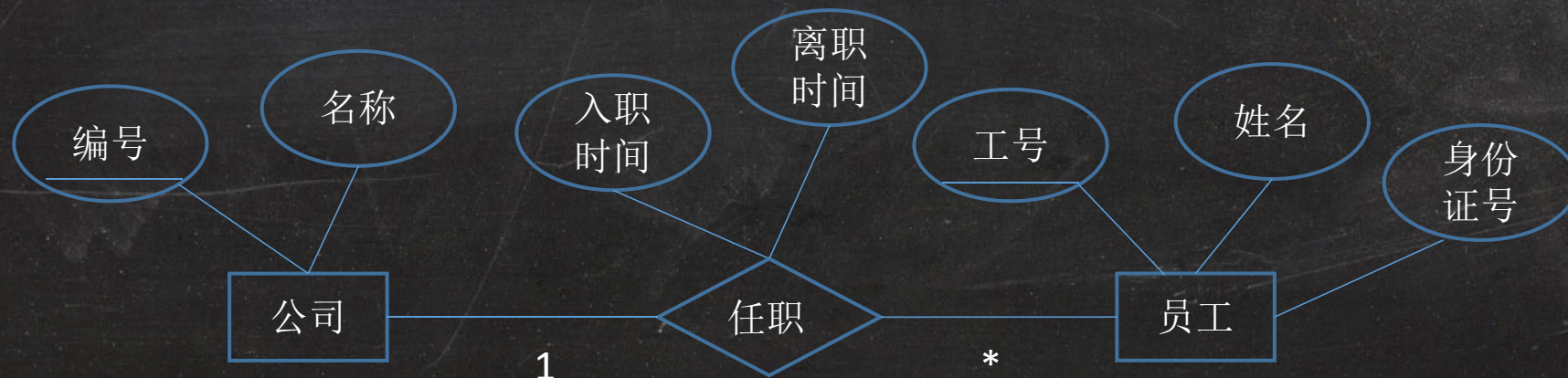
2、一对多联系的转换(1：*)：

(1) 将联系转换成一个独立的关系模式，关系模式的名称取联系的名称，关系模式的属性取该联系所关联的两个实体的码及联系的属性，关系的码是多方实体的码。

任职 (工号, 公司编号, 入职时间, 离职时间)

(2) 将联系归并到关联的两个实体的多方，给待归并的多方实体属性集中增加一方实体的码和该联系的属性即可，归并后的多方实体码保持不变。

员工 (工号, 姓名, 身份证号, 公司编号, 入职时间, 离职时间)



联系向关系模式的转换

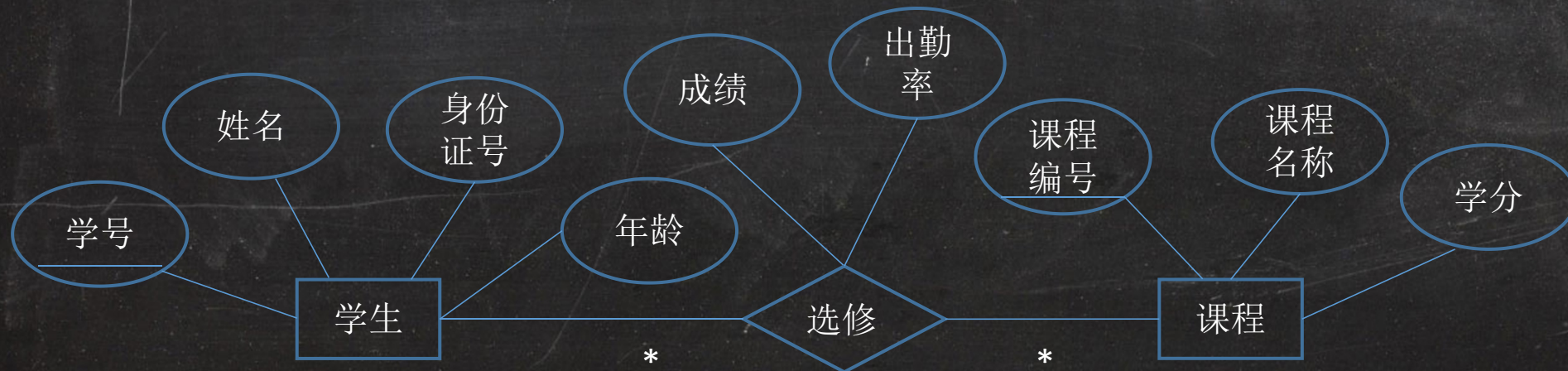
3、多对多联系的转换(*:*)：

- 多对多联系只能转换成一个独立的关系模式，关系模式的名称取联系的名称，关系模式的属性取该联系所关联的两个多方实体的码及联系的属性，关系的码是多方实体的码构成的属性组。

选修（学号，课程编号，成绩，出勤率）

- 注：三方联系的多对多对多(*:*:*)也是一样，只能转换成一个独立的关系模式，如：一个供应商可以给多个项目供应多种零件，一个项目可以使用多个供应商供应的多种零件。

供应关系（供应商编号，项目编号，零件号，数量）



14年第58、59题

在设计分E-R图阶段，人力部门定义的员工实体具有属性：员工号，姓名，性别和出生日期；教学部门定义的教师实体具有属性：教工号，姓名和职称。这种情况属于（ ），合并E-R图时，解决这一冲突的方法是（ ）。

(58) A.属性冲突 B.命名冲突 C.结构冲突 D.实体冲突

(59) A.员工和教师实体保持各自属性不变
B.员工实体中加入职称属性，删除教师实体
C.将教师实体所有属性并入员工实体，删除教师实体
D.将教师实体删除

14年第60、61题

某企业的E-R图中，职工实体的属性有：职工号、姓名、性别，出生日期，电话和所在部门，其中职工号为实体标识符，电话为多值属性，离退休职工所在部门为离退休办。在逻辑设计阶段，应将职工号和电话单独构造一个关系模式。该关系模式为（ ）；因为离退休职工不参与企业的绝大部分业务。应将这部分职工独立建立一个离退休职工关系模式，这种处理方式成为（ ）。

(60) A.1NF B.2NF C.3NF D.4NF

(61) A.水平分解 B.垂直分解 C.规范化 D.逆规范化

15年第28题

数据流图（DFD）的作用是（ ）。

- A.描述数据对象之间的关系
- C.说明将要出现的逻辑判定

- B.描述对数据的处理流程
- D.指明系统对外部事件的反应

15年第32~34题

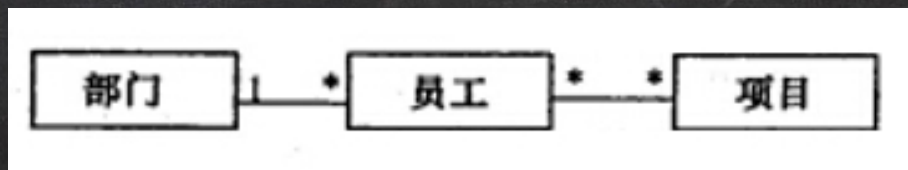
部门、员工和项目的关系模式及它们之间的E-R图如下所示，其中，关系模式中带实下划线的属性表示主键属性。

部门（部门代码，部门名称，电话）

员工（员工代码，姓名，部门代码，联系方式，薪资）

项目（项目编号，项目名称，承担任务）

若部门和员工关系进行自然连接运算，其结果集为（ ）元关系。由于员工和项目关系之间的联系类型为（ ），所以员工和项目之间的联系需要转换成一个独立的关系模式，该关系模式的主键是（ ）。



(32) A.5

B. 6

C. 7

D. 8

(33) A.1对1

B. 1对多

C. 多对1

D. 多对多

(34) A.（项目名称，员工代码）

B.（项目编号，员工代码）

C.（项目名称，部门代码）

D.（项目名称，承担任务）

15年第44题

数据字典中“数据项”的内容包括：名称、编号、取值范围、长度和（ ）。

A.处理频率

B.最大记录数

C.数据类型

D.数据流量

16年第29题

在某企业的信息综合管理系统设计阶段，员工实体在质量管理子系统被称为“质检员”，而在人事管理子系统被称为“员工”，这类冲突被称之为（ ）。

- A.语义冲突 B.命名冲突 C.属性冲突 D.结构冲突

16年第32题

关系规范化是在数据库设计的（ ）阶段进行。

- A.需求分析 B.概念设计 C.逻辑设计 D.物理设计

16年第50题

若系统使用频度最高的查询语句为

```
SELECT *  
FROM SC  
WHERE Sno = x AND Cno = y;                      // 其中，x, y 为变量
```

为使该查询语句的执行效率最高，应创建（ ）。

- A.Sno上的索引 B.Cno上的索引 C.Sno,Cno上的索引 D.SC上的视图SC_V (Sno,Cno)

16年第60、61题

数据库重构是指因为性能原因，对数据库中的某个表进行分解，再通过建立与原表同名的（ ）以保证查询该表的应用程序不变；通过修改更新原表的（ ）以保证外部程序对数据库的更新调用不变。

- (60)A.视图 B.索引 C.存储过程 D.触发器

- (61)A.视图 B.索引 C.存储过程 D.触发器

17年第38题

数据库概念结构设计阶段的工作步骤依次为（ ）。

- A.设计局部视图→抽象→修改重构消除冗余→合并取消冲突
- B.设计局部视图→抽象→合并取消冲突→修改重构消除冗余
- C.抽象→设计局部视图→修改重构消除冗余→合并取消冲突
- D.抽象→设计局部视图→合并取消冲突→修改重构消除冗余

17年第40、41题

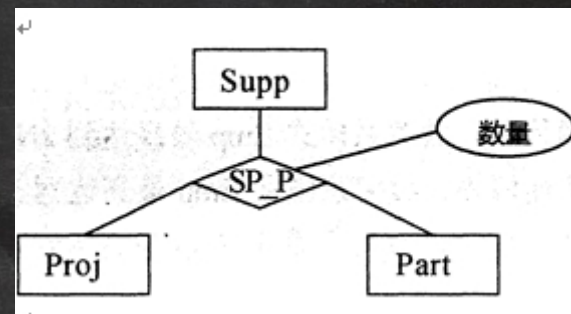
在某企业的工程项目管理数据库中供应商关系Supp、项目关系Proj 和零件关系Part 的E-R 模型和关系模式如下：

Supp (供应商号, 供应商名, 地址, 电话) //供应商号唯一标识Supp中的每一个元组

Proj (项目号, 项目名, 负责人, 电话) //项目号唯一标识 Proj 中的每一个元组

Part (零件号, 零件名) //零件号唯一标识 Part 中的每一个元组

其中，每个供应商可以为多个项目供应多种零件，每个项目可以由多个供应商供应多种零件，每种零件可以由多个供应商供应给多个项目。SP_P 的联系类型为（ ），（ ）。



- (40) A.*:*:* B.1:*:* C.1:1:* D.1:1:1

(41) A.不需要生成一个独立的关系模式

- B.需要生成一个独立的关系模式，该模式的主键为(项目号, 零件号, 数量)
- C.需要生成一个独立的关系模式，该模式的主键为(供应商号, 数量)
- D.需要生成一个独立的关系模式，该模式的主键为(供应商号, 项目号, 零件号)

17年第29题

数据库概念结构设计阶段是在（ ）的基础上，依照用户需求对信息进行分类、聚集和概括，建立概念模型。

- A.逻辑设计 B.需求分析 C.物理设计 D.运行维护

18年第64题

下列描述的因素中，不属于数据库物理设计应该考虑的是（ ）

- A. 事务的执行频度
- B. 使用频繁的查询操作
- C. 视图设计
- D. 索引设计

18年第65~67题

某数据库应用系统在运行过程中，发现随着数据量的不断增加，有部分查询业务和数据更新业务执行耗时越来越长。经分析，这些业务都与表 Table1 有关。假设 Table1 有 30 多个字段，分别为（Key，A1，A2，...Am，B1，B2，...Bn）。执行频度较高的查询业务都只用到 A1，A2，...Am 中的大部分属性，因此，DBA 决定将表 Table1 分解为 Table2（Key，A1，A2，...Am）和 Table3（Key，B1，B2，...Bn）。为了使所有对 Table1 的查询程序不必修改，应该（ ）；为了使对 Table1 的更新业务能正确执行，应该（ ）；这样实现了（ ）。

- (65) A. 修改所有对 Table1 的查询程序 B. 创建视图 Table1，为 Table2 和 Table3 的自然连接
- C. 只修改使用 A1，A2，...Am 中属性的程序 D. 只修改使用 B1，B2，...Bn 中属性的程序
- (66) A. 修改所有对 Table1 更新的事务程序
- B. 创建视图 Table1，为 Table2 和 Table3 的自然连接
- C. 只修改对 A1，A2，...Am 中属性更新的事务程序
- D. 只修改对 B1，B2，...Bn 中属性更新的事务程序
- (67) A. 数据的逻辑独立性 B. 数据的物理独立性
- C. 程序的逻辑独立性 D. 程序的物理独立性

18年第61题

需求分析阶段生成的文档中，用来描述企业中各项业务流程的是（ ）。

- A. 数据字典 B. 数据流图 C. E—R 图 D. 关系图

19年第61题

在设计关系模式时，有时为了提高数据操作的性能，会故意增加冗余数据，使得关系模式不满足3NF或BCNF，这种方法称之为反规范化，下列不属于反规范化手段的是（ ）。

- A. 合并模式
- B. 增加冗余属性
- C. 创建视图
- D. 增加派生属性

19年第62~64题

在索引改进中，一般的调整原则是：当（ ）是性能瓶颈时，则在关系上建立索引，当（ ）是性能瓶颈时，则考虑删除某些索引；管理人员经常会将有利于大多数数据查询的索引设为（ ）。

- A. 查询
- B. 更新
- C. 排序
- D. 分组计算
- A. 查询
- B. 更新
- C. 排序
- D. 分组计算
- A. 树索引
- B. 位图索引
- C. 散列索引
- D. 聚簇索引

19年第65、66题

在数据库系统运行中，经常会找出频繁执行的SQL语句进行优化。常见的优化策略有：尽可能减少多表查询或建立（ ）；用带（ ）的条件子句等价替换OR子句；只检索需要的属性列等。

- A. 视图
- B. 物化视图
- C. 外键约束
- D. 临时表
- A. IN
- B. EXISTS
- C. UNION
- D. AND

19年第67题

以下有关数据审计的叙述中，错误的是（ ）。

- A. 审计记录数据库资源和权限的使用情况
- B. 审计可以防止对数据库的非法修改
- C. 审计操作会影响系统性能
- D. 审计跟踪信息会扩大对存储空间的要求

20年第62题

ER图向关系模式转换时，实体标识符转换为关系的（ ）。

- A. 码
- B. 元组
- B. 记录
- D. 约束

21年第61题

在数据库设计中，下列步骤排序正确的选项是（ ）。

- ①需求分析 ②物理结构设计 ③概念结构设计 ④逻辑结构设计
A. ①②③④ B. ③①②④ C. ①④③② D. ①③④②

21年第62题

以下关于数据库设计的说法中，正确的是（ ）。

- A. 在逻辑结构设计阶段，规范化程度越高越好
- B. 逻辑结构设计的结果必须满足BCNF
- C. 在物理结构设计阶段，聚簇可提高特定属性的查询效率
- D. 在物理结构设计阶段，若选择B+树索引存取方法，关系上定义的索引数越多越好

21年第67题

以下关于数据库的重组和重构的说法中，正确的是（ ）。

- A. 数据库的重组修改了原设计的逻辑和物理结构
- B. 数据库的重构不修改原设计的逻辑和物理结构
- C. 数据库的重组是指按原设计要求重新安排存储位置、回收垃圾、减少指针链等，以提高系统性能
- D. 数据库的重构是指按原设计要求重新安排存储位置、回收垃圾、减少指针链

【22年第61题】

数据库设计的6个阶段依次是：需求分析，（ ），数据库实施，数据库运行和维护。

- A.逻辑结构设计、概念结构设计、物理结构设计
- B.概念结构设计、逻辑结构设计、物理结构设计
- C.概念结构设计、物理结构设计、逻辑结构设计
- D.物理结构设计、逻辑结构设计、概念结构设计

【22年第62题】

数据字典是（ ）。

- A.数据库中的字典
- B.查询数据的字典
- C.关于数据库中数据的描述
- D.不可更新的

【22年第63~64题】

某学生想要设计一个博客系统，在数据库的概念设计阶段提出了如下图所示的E-R模型，其中的联系类型为（ ），联系的属性为（ ）。

- (63) A.一元联系 B.二元联系 C.三元联系 D.四元联系
(64) A.发表 B.发表时间 C.用户 D.无

