
《数据库原理及应用》复习重点
第一章 数据库系统基本概念

要求、目标：

了解和掌握数据管理技术的发展阶段，数据描述的术语，数据抽象的四个级别，数据库管理系统的功能，数据库系统的组成。

一、数据管理技术的发展

1. 分为四个阶段：人工管理阶段、文件系统阶段、数据库阶段和高级数据库阶段。
2. 数据库阶段数据管理的特点：
 - 1) 采用数据模型表示复杂的数据结构。
 - 2) 有较高的数据独立性。
 - 3) 数据库系统为用户提供了方便的用户接口。
 - 4) 提供四方面的数据控制功能：数据库的恢复、数据库的并发控制、数据的完整性、数据安全性。
 - 5) 增加了系统的灵活性。
3. 数据库（DB）：是长期存储在计算机内、有组织的、统一管理的相关数据的集合。
4. 数据库管理系统（DBMS）：是位于用户与操作系统之间的一层数据管理软件，它为用户或应用程序提供访问DB的方法，包括DB的建立、查询、更新及各种数据控制。
5. 数据库系统（DBS）：是实现有组织地、动态地存储大量关联数据、方便多用户访问的计算机硬件、软件和数据资源组成的系统，即它是采用数据库技术的计算机系统。

二、数据描述

1. 分为三个阶段：概念设计、逻辑设计和物理设计。
2. 概念设计中的术语：
 - 1) 实体：客观存在，可以相互区别的事物称为实体。
 - 2) 实体集：性质相同的同类实体的集合。
 - 3) 属性：实体有很多特性，每一个特性称为属性。
 - 4) 实体标识符（关键码或键）：能惟一标识实体的属性或属性集。以上概念均有类型和值之分。
3. 逻辑设计中的术语：
 - 1) 字段（数据项）：标记实体属性的命名单位称为字段或数据项。
 - 2) 记录：字段的有序集合。
 - 3) 文件：同一类记录的集合。
 - 4) 关键码：能惟一标识文件中每个记录的字段或字段集。以上概念均有类型和值之分。
4. 概念设计和逻辑设计中术语的对应关系：

概念设计		逻辑设计
实体	—	记录
属性	—	字段（数据项）
实体集	—	文件
实体标识符	—	关键码
5. 实体之间联系的元数：与一个联系有关的实体集个数。常用二元联系。二元联系的类型有三种：一对一联系、一对多联系、多对多联系。
6. 一对一联系：如果实体集E1中每个实体至多和实体集E2中的一个实体有联系，反之亦然，那么实体集E1和E2的联系称为“一对一联系”，记为“1: 1”。
7. 一对多联系：如果实体集E1中每个实体与实体集E2中任意个（零个或多个）实体间有联系，而E2中每个实体至多和E1中一个实体有联系，那么称E1对E2的联系是“一对多联系”，记为“1: N”。
8. 多对多联系：如果实体集E1中每个实体可以与实体集E2中任意个（零个或多个）实体间有联系，反之亦然，那么称E1和E2的联系是“多对多联系”，记为“M: N”。

三、数据抽象的级别

1. 数据模型：描述数据库的结构和定义，对现实世界的数据进行抽象。
2. 从现实世界的信息到数据库存储的数据以及用户使用的数据是一个逐步抽象过程，根据数据抽象的级别定义了四种模型：概念模型、逻辑模型、外部模型和内部模型。
3. 概念模型：表达用户需求观点的数据全局逻辑结构的模型。
4. 逻辑模型：表达计算机实现观点的DB全局逻辑结构的模型。

-
5. 外部模型：表达用户使用观点的DB局部逻辑结构的模型。
 6. 内部模型：表达DB物理结构的模型。
 7. 数据抽象的过程、即数据库设计的过程具体步骤：
 - 1) 根据用户需求，设计数据库的概念模型；
 - 2) 根据转换规则，把概念模型转换成数据库的逻辑模型；
 - 3) 根据用户的业务特点，设计不同的外部模型，给程序员使用；
 - 4) 数据库实现时，要根据逻辑模型设计其内部模型。通常分为概念设计、逻辑设计（2和3步）和物理设计三个阶段。
 8. 常用的概念模型是实体联系（ER）模型，ER模型主要用ER图来表示。
 9. 逻辑模型的分类：层次模型、网状模型、关系模型等。
 10. 层次模型：用树型（层次）结构表示实体及实体间联系的数据模型。
 11. 1969年，美国IBM公司的IMS系统是典型的层次模型系统。
 12. 网状模型：用有向图结构表示实体及实体间联系的数据模型。
 13. 1969年，CODASYL组织提出DBTG报告中的数据模型是网状模型的主要代表。
 14. 关系模型：是由若干个关系模式组成的集合。关系模式即记录类型，它的实例称为关系，每个关系实际上是一张二维表格。
 15. 1970年，美国IBM公司的E.F.Codd连续发表论文，提出关系模型，奠定了关系数据库的理论基础。关系数据库是目前的主流数据库。
 16. 外部模型中的模式称为视图。
 17. 三级模式：

从用户（或应用程序）到数据库之间，DB的数据结构描述有三个层次：

 - 1) 外模式：用户与数据库系统的接口，是用户用到的那部分数据的描述。外模式由若干个记录类型组成。
 - 2) 逻辑模式：是数据库中全部数据的整体逻辑结构的描述。
 - 3) 内模式：是数据库在物理存储方面的描述。注意：外模式是逻辑模式的子集。
 18. 两级映像：
 - 1) 外模式/逻辑模式映像：存在于外模式和逻辑模式之间，用于定义外模式和逻辑模式之间的对应性。
 - 2) 逻辑模式/内模式映像：存在于逻辑模式和内模式之间，用于定义逻辑模式和内模式之间的对应性。
 19. 数据库系统的三级模式、两级映像结构使数据库系统达到了高度的数据独立性。
 20. 数据独立性：是指应用程序与数据库的数据结构之间相互独立，在修改数据结构时，尽可能不修改应用程序。分为逻辑数据独立性和物理数据独立性。
 21. 逻辑数据独立性：如果数据库的逻辑模式要修改，那么只要对外模式逻辑模式映像作相应的修改，可以使外模式和应用程序尽可能保持不变。这样就认为数据库达到了逻辑数据独立性。
 22. 物理数据独立性：如果数据库的内模式要修改，即数据库的物理结构有所变化，那么只要对逻辑模式/内模式映像作相应的修改，可以使逻辑模式尽可能保持不变。也就是对内模式的修改尽量不影响逻辑模式，当然对外模式和应用程序的影响更小，这样就认为数据库达到了物理数据独立性。

四、数据库管理系统 DBMS

1. DBMS的主要功能：数据库的定义功能（DBMS提供DDL定义数据库的三级模式、两级映像等）、数据库的操纵功能（DBMS提供DML实现对数据的操作，基本的数据操作有检索和更新两类）、数据库的保护功能、数据库的维护功能、数据字典。

五、数据库系统 DBS

1. DBS的组成：是数据库、硬件、软件和数据库管理员的集合体。
2. 软件包括DBMS、OS、各种主语言和应用开发支撑软件等程序。其中，DBMS是DBS的核心软件，要在OS支持下才能工作。
3. 数据库管理员（DBA）：是控制数据整体结构的一组人员，负责DBS的正常运行，承担创建、监控和维护数据库结构的责任。

第一章 复习题

一、单项选择题

1. 在数据库系统中，当数据库的内模式发生改变时，应用程序也可以不变。这是（A）
A) 物理数据独立性 B) 逻辑数据独立性

-
- C) 位置数据独立性 D) 存储数据独立性
2. 在下面列出的数据模型中, 哪一个是概念数据模型 (D)
A) 关系模型 B) 层次模型 C) 网状模型 D) 实体-联系模型
3. 下面列出的数据管理技术发展阶段中, 哪个(些)阶段数据不能保存在计算机中? (A)
I. 人工管理阶段
II. 文件系统阶段
III. 数据库阶段
A) 只有 I B) 只有 II C) I 和 II D) II 和 III
4. 用二维表结构表示实体以及实体间联系的数据模型称为 (C)
A) 网状模型 B) 层次模型 C) 关系模型 D) 实体-联系模型
5. 下面列出的条目中, 哪些是数据库技术的主要特点 (A)
I. 数据共享度高
II. 数据的冗余小
III. 较高的数据独立性
IV. 程序的标准化
A) I、II 和 III B) I 和 II C) I、II 和 IV D) 都是
6. 数据库管理系统 DBMS 中用来定义逻辑模式、内模式和外模式的语言是 (C)
A) DML B) C C) DDL D) Basic
7. 层次模型的典型代表是 (A) 数据库管理系统。
A) IMS B) IBM C) DBTG D) ATM
8. 负责数据库系统的正常运行, 承担创建、监控和维护数据库结构责任的是 (C)
A) 应用程序员 B) 终端用户
C) 数据库管理员 D) 数据库管理系统的软件设计员
9. 在以下各条叙述中, 正确的叙述有几条 (B)
1) 数据库避免了一切数据重复
2) 数据库减少了数据冗余
3) 数据库中, 如果逻辑模式改变, 则需将与其有关的外模式做相应改变, 否则应用程序需改写
4) 数据库中的内模式如有改变, 逻辑模式可以不变
A) 1 B) 2 C) 3 D) 4
10. 在三级模式之间引入两级映像, 其主要功能之一是 (A)
A) 使数据与程序具有较高的独立性
B) 使系统具有较高的通道能力
C) 保持数据与程序的一致性
D) 提高存储空间的利用率
11. 视图对应于数据库系统三级模式结构中的 (A)
A) 外模式 B) 逻辑模式 C) 内模式 D) 都不对应
12. 对数据库中数据可以进行查询、插入、删除、修改, 这是因为数据库管理系统提供了 (B)
A) 数据库定义功能 B) 数据库操纵功能
C) 数据库维护功能 D) 数据库控制功能
13. 1970 年, 美国 IBM 公司研究员 E.F.Codd 提出了数据库的 (C)
A) 层次模型 B) 网状模型 C) 关系模型 D) 实体联系模型
14. 具有数据冗余度小、数据共享、以及较高数据独立性等特征的系统是 (B)
A) 文件系统 B) 数据库系统 C) 管理系统 D) 高级程序
15. 在概念设计中的事物称为 (A)
A) 实体 B) 记录 C) 对象 D) 结点
16. 层次模型表达实体及实体之间联系用的数据结构是 (D)
A) 网状 B) 表 C) 有向图 D) 树型
17. DB 是 (A)
A) 数据库 B) 数据库管理系统
C) 数据处理系统 D) 数据库系统
18. DBMS 是 (B)
A) 数据库 B) 数据库管理系统
C) 数据处理系统 D) 数据库系统

19. 已知在一个工厂中有多个车间，每一个车间有多名职工，工厂的产品要经过多个车间的多道工序加工。具体来说，一个产品要经过多个工人加工，一位工人要加工多个产品。问：工厂与车间之间属于（A）联系，车间与工人之间属于（A）联系，工人与产品之间属于（D）联系。
A）一对多 B）一对一 C）未知 D）多对多
20. （D）是存储在计算机内的有结构的数据集合。
A）网络系统 B）数据库系统 C）操作系统 D）数据库
21. 数据库系统的核心是（D）
A）编译系统 B）数据库 C）操作系统 D）数据库管理系统
22. 数据库（DB）、数据库系统（DBS）和数据库管理系统（DBMS）三者之间的关系是（A）
A）DBS包括DB和DBMS B）DBM包括DB和DBS
C）DB包括DBS和DBMS D）DBS就是DB 也就是DBMS
23. 数据库管理系统（DBMS）是（B）
A）一组硬件 B）一组软件 C）既有硬件，也有软件
24. 在DBS中，DBMS和OS之间关系是（B）
A）相互调用 B）DBMS调用OS
C）OS调用DBMS D）并发运行
25. 数据库系统是在（A）的基础上发展起来的。
A）文件系统 B）应用程序系统
C）数据库管理系统 D）编译系统
26. 三级模式间存在两种映射，它们是（A）
A）逻辑模式与外模式间，逻辑模式与内模式间
B）外模式与内模式间，外模式与逻辑模式间
C）内模式与外模式间，内模式与逻辑模式间
27. 表达用户使用观点的数据库局部逻辑结构的模型是（C）
A）概念模型 B）逻辑模型
C）外部模型 D）内部模型

二、填空题

1. 数据管理技术的发展经历了如下四个阶段：人工管理阶段、文件系统阶段、数据库阶段和高级数据库阶段。
2. 用二维表结构表示的实体及实体间联系的数据模型称为关系模型。
3. 两个实体集之间的联系有三种，分别是一对一联系，一对多联系和多对多联系。
4. 如果实体集E1中每个实体至多和实体集E2中的一个实体有联系，反之亦然，那么实体集E1和E2的联系称为一对一联系。
5. 数据库系统的三级模式、两级映像结构使数据库系统达到了高度的数据独立性。

第二章 数据库设计和ER模型

要求、目标：

了解和掌握数据库应用系统设计的全过程，掌握ER模型和关系模型的基本概念，掌握概念设计ER模型的设计方法，掌握逻辑设计ER模型向关系模型转换的方法。

一、数据库系统生存期

1. 数据库系统生存期：数据库应用系统从开始规划、设计、实现、维护到最后被新的系统取代而停止使用的整个期间。
2. 数据库系统生存期分七个阶段：规划、需求分析、概念设计、逻辑设计、物理设计、实现、运行维护。
3. 规划阶段三个步骤：系统调查、可行性分析、确定数据库系统总目标。
4. 需求分析阶段：主要任务是系统分析员和用户双方共同收集数据库系统所需要的信息内容和用户对处理的需求，并以需求说明书的形式确定下来。
5. 概念设计阶段：产生反映用户单位信息需求的概念模型。与硬件和DBMS无关。
6. 逻辑设计阶段：将概念模型转换成DBMS能处理的逻辑模型。外模型也将在此阶段完成。
7. 物理设计阶段：对于给定的基本数据模型选取一个最适合应用环境的物理结构的过程。数据库的物理结构主要指数据库的存储记录格式、存储记录安排和存取方法。
8. 数据库的实现：包括定义数据库结构、数据装载、编制与调试应用程序、数据库试运行。

二、ER模型的基本概念ER模型的基本元素是：实体、联系和属性。

2. 实体：是一个数据对象，指应用中可以区别的客观存在的事物。

实体集：是指同一类实体构成的集合。

实体类型：是对实体集中实体的定义。

一般将实体、实体集、实体类型统称为实体。

3. 联系：表示一个或多个实体之间的关联关系。

联系集：是指同一类联系构成的集合。

联系类型：是对联系集中联系的定义。

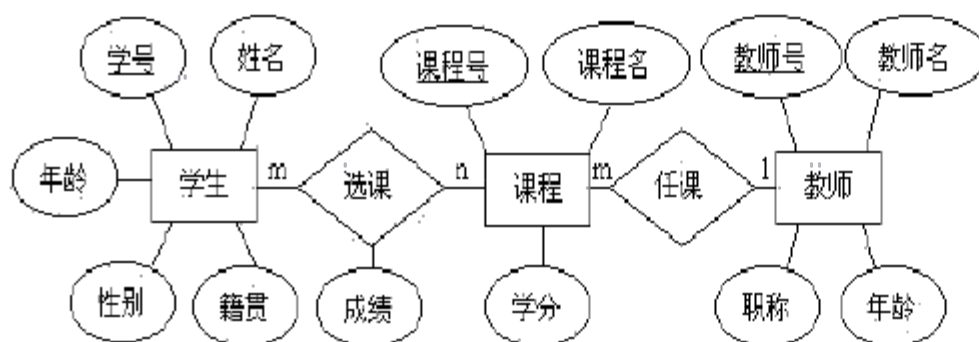
一般将联系、联系集、联系类型统称为联系。

4. 同一个实体集内部实体之间的联系，称为一元联系；两个不同实体集实体之间的联系，称为二元联系，以此类推。

5. 属性：实体的某一特性称为属性。在一个实体中，能够惟一标识实体的属性或属性集称为实体标识符。

6. ER 模型中，方框表示实体、菱形框表示联系、椭圆形框表示属性、实体与联系、实体与其属性、联系与其属性之间用直线连接。实体标识符下画横线。联系的类型要在直线上标注。注意：联系也有可能存在属性，但联系本身没有标识符。

例：假设一个学生可选多门课程，而一门课程又有多个学生选修，一个教师可讲多门课程，一门课程至多只有一个教师讲授。ER 图如下：



7. 概念设计三个步骤：设计局部ER 模型、设计全局ER 模型和全局 ER 模型的优化。

三、关系模型的基本概念

1. 关系模型的定义：用二维表格表示实体集，用关键码表示实体之间联系的数据模型。

2. 在关系模型中，字段称为属性，字段值称为属性值，记录类型称为关系模式。记录称为元组，元组的集合称为关系或实例。有时习惯称关系为表或表格，元组为行，属性为列。关系中属性个数称为元数，元组个数称为基数。

3. 关键码（简称键）：由一个或多个属性组成。

4. 超键：在关系中能惟一标识元组的属性集称为关系模式的超键。

5. 候选键：不含有多余属性的超键。

6. 主键：用户选作元组标识的候选键。一般如不加说明，键是指主键。

7. 外键：如果模式R 中属性集K 是其他模式的主键，那么K 在模式R 中称为外键。

8. 值域：关系中每一个属性都有一个取值范围，称为属性的值域。每一个属性对应一个值域，不同的属性可对应于同一值域。

9. 关系的定义：关系是一个属性数目相同的元组的集合。

10. 关系的性质：关系是一种规范化了的二维表格。

1) 关系中每一个属性值都是不可分解的；

2) 关系中不允许出现重复元组；

3) 关系没有行序；

4) 元组中的属性在理论上也是无序的，但使用时按习惯考虑列的顺序。

11. 关系数据库中的数据与更新操作必须遵循三类完整性规则：实体完整性规则、参照完整性规则、用户定义的完整性规则。

12. 实体完整性规则：要求关系中元组在组成主键的属性上不能有空值。

13. 参照完整性规则：如果属性集K 是关系模式R1 的主键，K 也是关系模式R2 的外键，那么在R2 关系中，K 的取值只允许两种可能，或者为空值，或者等于R1 关系中某个主键值。这条规则的实质是“不允许引用不存在的实体”。其中，R1 称为参照关系；R2 称为依赖关系。

注意：这条规则在具体使用时，有三点变通：

1) 外键和相应的主键可以不同名，只要定义在相同值域上即可；

2) R1 和 R2 可以是同一个关系模式，此时表示了同一个关系中不同元组之间的联系；

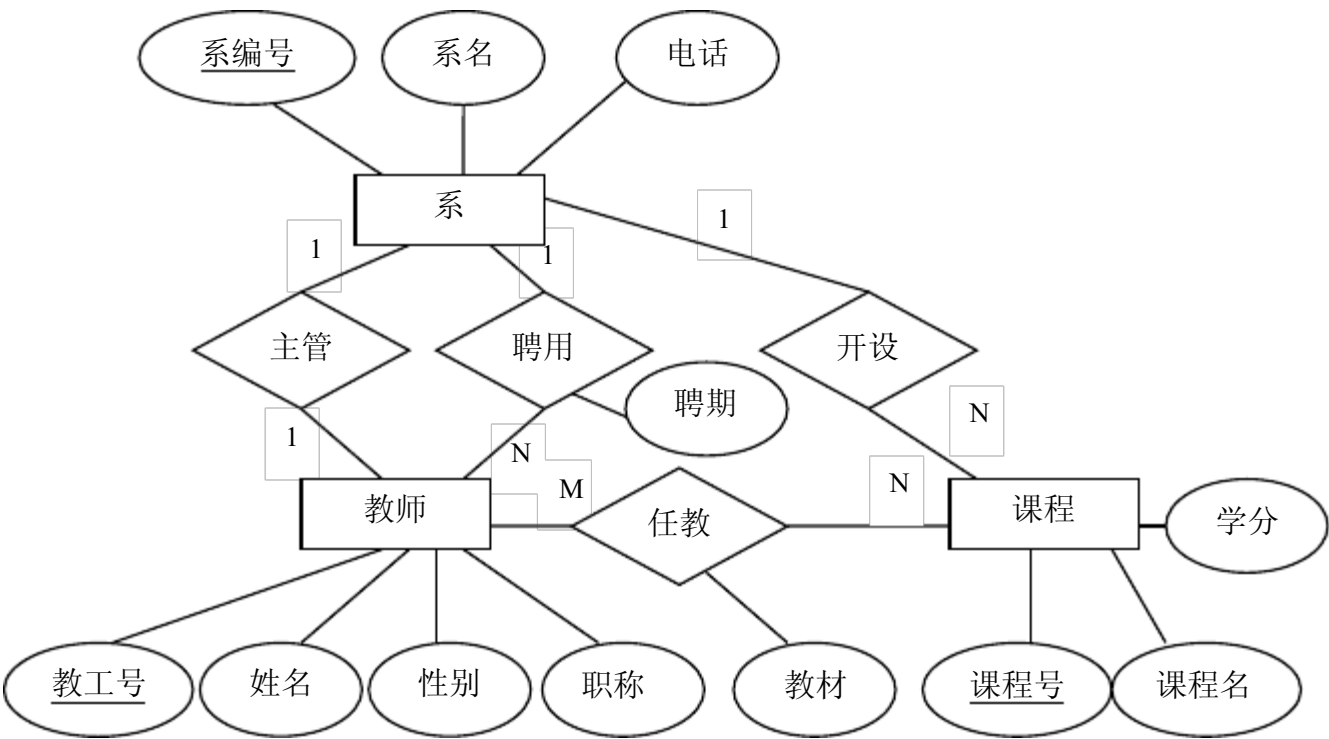
3) 外键值是否允许空，应视具体问题而定。

14. 用户定义的完整性规则：用户针对具体的数据约束，设置的完整性规则，由系统来检验实施。

四、ER 模型到关系模型的转换

1. ER 图转换成关系模式集的算法：

- 1) 实体类型的转换：将每个实体类型转换成一个关系模式，实体的属性即为关系模式的属性，实体标识符即为关系模式的键。
- 2) 联系类型的转换：主要掌握二元联系类型的转换。
- a) 若实体间联系是 1: 1，可以在两个实体类型转换成的关系模式中任意一个关系模式的属性中加入另一个关系模式的键（作为外键）和联系类型的属性。
 - b) 若实体间联系是 1: N，则在 N 端实体类型转换成的关系模式中加入 1 端实体类型的键（作为外键）和联系类型的属性。
 - c) 若实体间联系是 M: N，则将联系类型也转换成关系模式，其属性为两端实体类型的键（作为外键）加上联系类型的属性，而键为两端实体键的组合。
- 例：下面是教学管理的一个可能的 ER 图。图中，有三个实体类型：系、教师和课程；有四个联系类型：主管、聘用、开设和任教。根据转换算法，把该图转换成关系模式集的步骤如下：



教学管理的ER图

- 第一步：把三个实体类型转换成三个关系模式（注意关系模式的表示方法）：
- 系（系编号，系名，电话）
 - 教师（教工号，姓名，性别，职称）
 - 课程（课程号，课程名，学分）
- 第二步：对于 1: 1 联系，可以在“系”模式中加入教工号（教工号为外键）。
- 对于 1: N 联系“聘用”，可以在“教师”模式中加入系编号和聘期两个属性（系编号为外键）；
- 对于 1: N 联系“开设”，可以在“课程”模式中加入系编号（系编号为外键）。这样第一步得到的三个模式改变成如下形式（注意：外键通常使用波浪线表示）：
- 系（系编号，系名，电话，主管人的教工号）
 - 教师（教工号，姓名，性别，职称，系编号，聘期）
 - 课程（课程号，课程名，学分，系编号）
- 第三步：对于 M: N 联系“任教”，则生成一个新的关系模式：
- 任教（教工号，课程号，教材）
- 这样，转换成的四个关系模式如下：
- 系（系编号，系名，电话，主管人的教工号）
 - 教师（教工号，姓名，性别，职称，系编号，聘期）
 - 课程（课程号，课程名，学分，系编号）
 - 任教（教工号，课程号，教材）

2. 采用 ER 模型的逻辑设计步骤

- 1) 导出初始关系模式集：把概念设计的结果（即全局ER模型）转换成初始关系模式集。
- 2) 规范化处理
- 3) 模式评价
- 4) 模式修正
- 5) 设计外模式

第二章复习题

一、单项选择题

1. 下面对于关系的叙述中，哪个是不正确的？（C）
 - A) 关系中的每个属性是不可分解的
 - B) 在关系中元组的顺序是无关紧要的
 - C) 任意的一个二维表都是一个关系
 - D) 每一个关系只有一种记录类型
2. 设属性A是关系R的主键，则属性A不能取空值。这是（A）
 - A) 实体完整性规则
 - B) 参照完整性规则
 - C) 用户定义完整性规则
 - D) 域完整性规则
3. 对关系数据库来讲，下面（C）说法是错误的
 - A) 外键和相应的主键可以不同名，只要定义在相同值域上即可
 - B) 不同的属性可对应于同一值域
 - C) 行的顺序可以任意交换，但列的顺序不能任意交换
 - D) 关系中的任意两个元组不能完全相同
4. 数据库的物理设计是为一个给定的逻辑结构选取一个适合应用环境的（B）的过程，包括确定数据库在物理设备上的存储结构和存取方法。
 - A) 逻辑结构
 - B) 物理结构
 - C) 概念结构
 - D) 层次结构
5. 在关系中，能惟一标识元组的属性集称为（B）。
 - A) 外键
 - B) 超键
 - C) 域
 - D) 元组
6. 在数据库逻辑设计中，当将E-R图转换为关系模式时，下面的做法（B）是不正确的
 - A) 一个实体类型转换为一个关系模式
 - B) 一个联系类型转换为一个关系模式
 - C) 由实体类型转换成的关系模式的主键是该实体类型的主键
7. 数据库设计的概念设计阶段，表示概念结构的常用方法和描述工具是（D）
 - A) 层次分析法和层次结构图
 - B) 数据流程分析法和数据流程
 - C) 结构分析法和模块结构图
 - D) 实体—联系方法和E—R图
8. 关系数据库中，实现表与表之间的联系是通过（B）
 - A) 实体完整性规则
 - B) 参照完整性规则
 - C) 用户定义的完整性
 - D) 值域
9. 有两个实体集，并且它们之间存在着一个M N联系，那么按照E—R模型转换成关系数据库的规则，这个E—R结构转换成表的个数为（C）
 - A) 1
 - B) 2
 - C) 3
 - D) 4
10. 把E-R图转换为关系模型的过程，属于数据库设计的（B）
 - A) 概念设计
 - B) 逻辑设计
 - C) 需求分析
 - D) 物理设计
11. 在数据库设计的（D）阶段，进行应用程序的编制和调试。
 - A) 概念设计
 - B) 逻辑设计
 - C) 物理设计
 - D) 实现

12.

EMP

雇员号	雇员名	部门号	工资
001	张山	02	2000
010	王宏达	01	1200
056	马林生	02	1000
101	赵敏	04	1500

DEPT

部门号	部门名	地址
01	业务部	1号楼

02	销售部	2 号楼
03	服务部	3 号楼
04	财务部	4 号楼

在雇员信息表关系EMP中，哪个属性是外键？（C）

A) 雇员号 B) 雇员名 C) 部门号 D) 工资

二、填空题

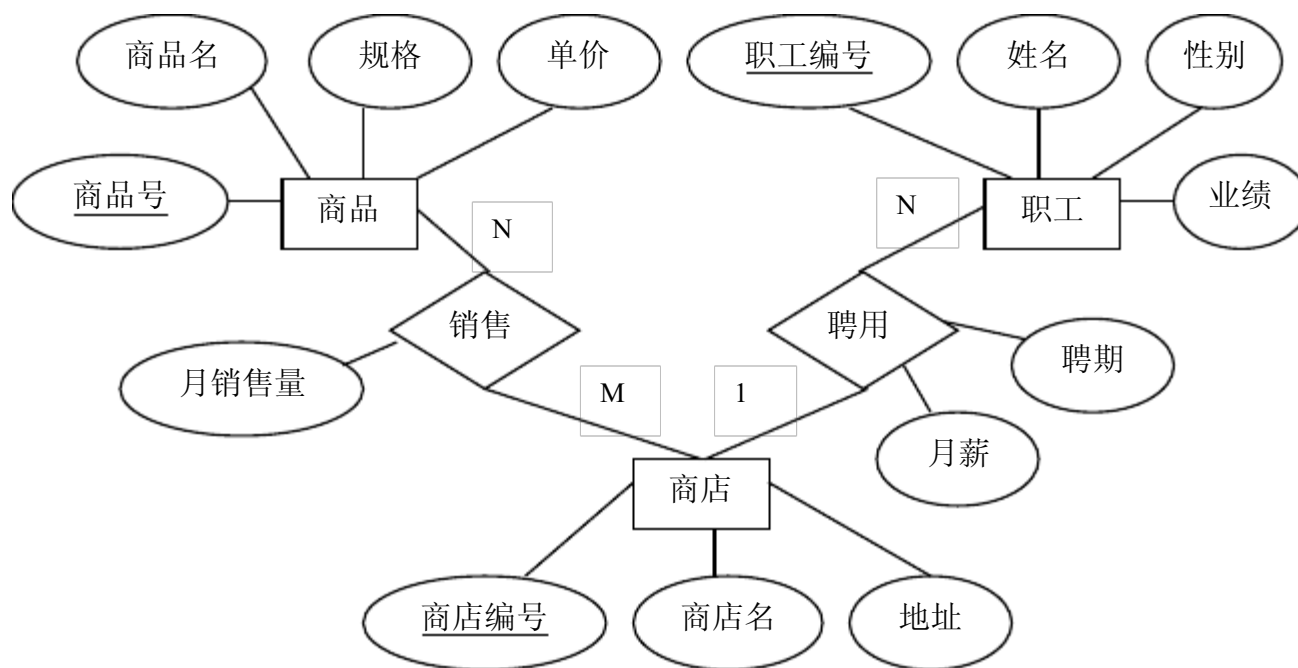
1. 表示实体及其联系的方法为 ER图，基本图素包括方框、菱形框和椭圆形框。习惯上实体用方框表示，实体的属性用椭圆形框表示，联系用菱形框表示。
2. 在关系模型中，二维表的列称为属性，二维表的行称为元组。
3. 假设班级和班长之间的联系是一对一联系，则班级和学生之间是一对多联系；另外，学生和课程之间的联系是多对多联系。
4. 关系模型的三种完整性规则是实体完整性规则、参照完整性规则和用户定义的完整性规则。
5. 已知系（系编号，系名称，系主任，电话，地点）和学生（学号，姓名，性别，入学日期，专业，系编号）两个关系，系关系的主键是系编号，学生关系的主键是学号，外键是系编号。
6. 数据库系统生存期分七个阶段：规划、需求分析、概念设计、逻辑设计、物理设计、实现、运行维护。
7. 关系中属性个数称为元数，元组个数称为基数。

三、综合题

1. 设某商业集团数据库中有三个实体集。一是“商店”实体集，属性有商店编号、商店名、地址等；二是“商品”实体集，属性有商品号、商品名、规格、单价等；三是“职工”实体集，属性有职工编号、姓名、性别、业绩等。

商店与商品间存在“销售”联系，每个商店可销售多种商品，每种商品也可放在多个商店销售，每个商店每销售一种商品，有月销售量；商店与职工间存在着“聘用”联系，每个商店有许多职工，每个职工只能在一个商店工作，商店聘用职工有聘期和月薪。

试画出 ER图，并在图上注明属性、联系的类型。再转换成关系模式集，并指出每个关系模式的主键和外键。



关系模式集：

商店（商店编号，商店名，地址）

商品（商品号，商品名，规格，单价）

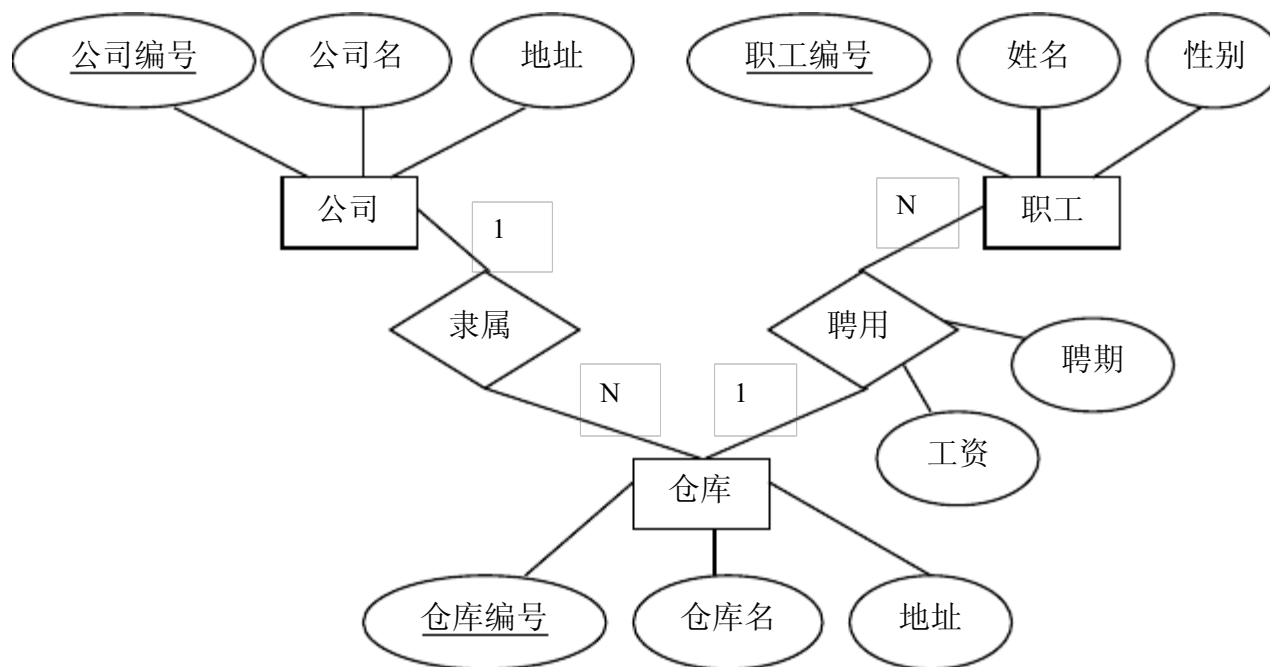
职工（职工编号，姓名，性别，业绩，商店编号，月薪，聘期）

销售（商店编号，商品编号，月销售量）

2. 设某商业集团数据库中有三个实体集。一是“公司”实体集，属性有公司编号、公司名、地址等；二是“仓库”实体集，属性有仓库编号、仓库名、地址等；三是“职工”实体集，属性有职工编号、姓名、性别等。

公司与仓库间存在“隶属”联系，每个公司管辖若干仓库，每个仓库只能属于一个公司管辖；仓库与职工间存在“聘用”联系，每个仓库可聘用多个职工，每个职工只能在一个仓库工作，仓库聘用职工有聘期和工资。

试画出 ER图，并在图上注明属性、联系的类型。再转换成关系模式集，并指出每个关系模式的主键和外键。



关系模式集：

公司（公司编号，公司名，地址）

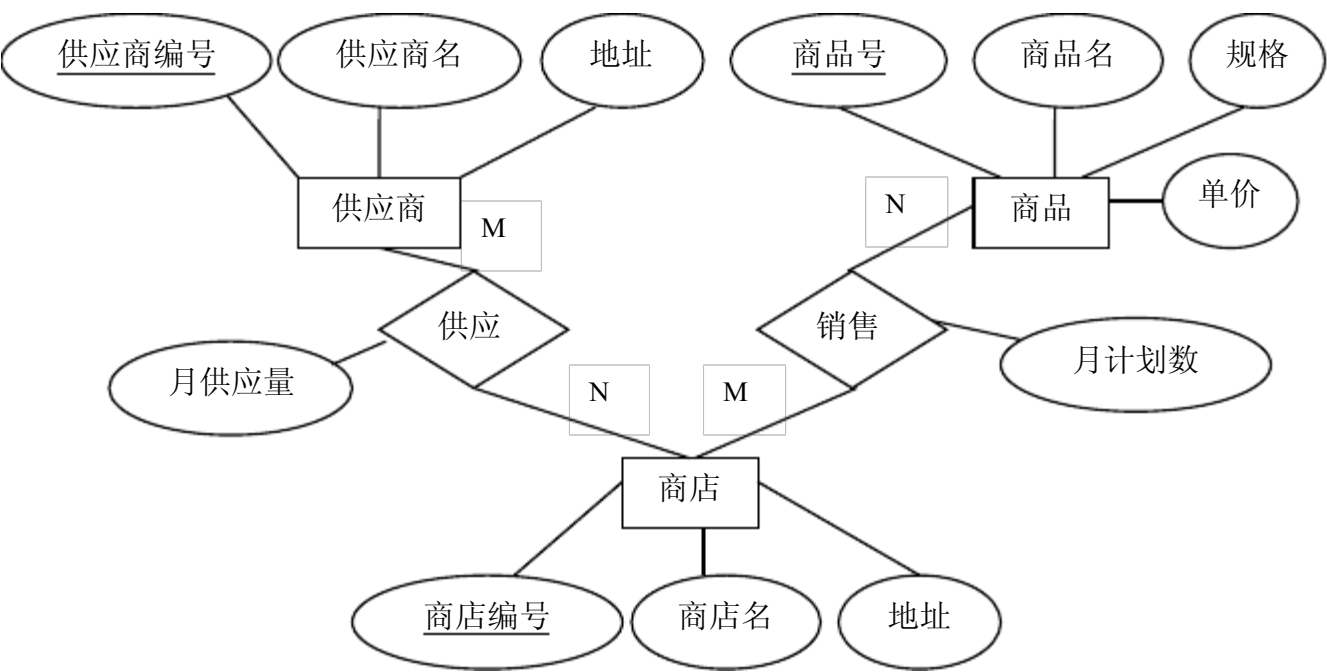
仓库（仓库编号，仓库名，地址，公司编号）

职工（职工编号，姓名，性别，仓库编号，聘期，工资）

3. 设某商业集团数据库中有三个实体集。一是“商品”实体集，属性有商品号、商品名、规格、单价等；二是“商店”实体集，属性有商店编号、商店名、地址等；三是“供应商”实体集，属性有供应商编号、供应商名、地址等。

供应商与商品间存在“供应”联系，每个供应商可供应多种商品，每种商品可向多个供应商订购，供应商供应商品有月供应量；商店与商品间存在“销售”联系，每个商店可销售多种商品，每种商品可在多个商店销售，商店销售商品有月计划数。

试画出 ER图，并在图上注明属性、联系的类型。再转换成关系模式集，并指出每个关系模式的主键和外键。



关系模式集：
 供应商（供应商编号，供应商名，地址）
 商店（商店编号，商店名，地址）
 商品（商品号，商品名，规格，单价）
 供应（供应商编号，商店编号，月供应量）
 销售（商店编号，商品号，月计划数）

第三章 关系模式设计理论

- 要求、目标：
- 了解关系数据库规范化理论及其在数据库设计中的作用，重点是函数依赖和范式，要求掌握这些概念并能运用它们来进行模式分解。
- 一、关系模式的设计准则
1. 数据冗余：同一个数据在系统中多次重复出现。
 2. 关系模式设计不当引起的异常问题：数据冗余、操作异常（包括修改异常、插入异常和删除异常）
 3. 关系模式的非形式化设计准则
 - 1) 关系模式的设计应尽可能只包含有直接联系的属性，不要包含有间接联系的属性。也就是，每个关系模式应只对应于一个实体类型或一个联系类型。
 - 2) 关系模式的设计应尽可能使得相应关系中不出现插入异常、删除和修改等操作异常现象。
 - 3) 关系模式的设计应尽可能使得相应关系中避免放置经常为空值的属性。
 - 4) 关系模式的设计应尽可能使得关系的等值连接在主键和外键的属性上进行，并且保证以后不会生成额外的元组。
 4. 习惯使用的一些符号：
 - 1) 英文字母表首部的大写字母“ A, B, C, ...”表示单个的属性。
 - 2) 英文字母表尾部的大写字母“ ..., U, V, W, X, Y, Z”表示属性集。
 - 3) 大写字母 R 表示关系模式，小写字母 r 表示其关系。
 - 4) 关系模式的简化表示方法：R (A, B, C, ...) 或 R (ABC...)

5) 属性集 X 和 Y 的并集简写为 XY。

二、函数依赖

1. 函数依赖 (FD) 的定义: 设有关系模式 $R(U)$, X 和 Y 是属性集 U 的子集, 函数依赖是形成 $X \rightarrow Y$ 的一个命题, 只要 r 是 R 的当前关系, 对 r 中任意两个元组 t 和 s, 都有 $t[X]=s[X]$ 蕴涵 $t[Y]=s[Y]$, 那么称 FD $X \rightarrow Y$ 在关系模式 $R(U)$ 中成立。

说明: 1) $t[X]$ 表示元组 t 在属性集 X 上的值, 其余类同。

2) $X \rightarrow Y$ 读作“X 函数决定 Y”或“Y 函数依赖于 X”。

3) FD 是对关系模式 R 的一切可能的关系 r 定义的。对于当前关系 r 的任意两个元组, 如果 X 值相同, 则要求 Y 值也相同, 即有一个 X 值就有一个 Y 值与之对应, 或者说 Y 值由 X 值决定。

例: 设关系模式 $R(ABCD)$, 在 R 的关系中, 属性值间有这样的联系: A 值与 B 值有一对多联系; C 值与 D 值之间有一对一联系。试根据这些规则写出相应的函数依赖。

$B \rightarrow A \quad C \rightarrow D \quad D \rightarrow C$

2. 如果 $X \rightarrow Y$ 和 $Y \rightarrow X$ 同时成立, 则可记为: $X \leftrightarrow Y$

3. FD 的逻辑蕴涵: 设 F 是在关系模式 R 上成立的函数依赖的集合, $X \rightarrow Y$ 是一个函数依赖。如果对于 R 的每个满足 F 的关系 r 也满足 $X \rightarrow Y$, 那么称 F 逻辑蕴涵 $X \rightarrow Y$, 记为 $F \models X \rightarrow Y$ 。

4. 设 F 是函数依赖集, 被 F 逻辑蕴涵的函数依赖全体构成的集合, 称为函数依赖集 F 的闭包, 记为 F^+ 。即 $F^+ = \{X \rightarrow Y \mid F \models X \rightarrow Y\}$

5. FD 的推理规则 (Armstrong 公理)

设 U 是关系模式 R 的属性集, F 是 R 上成立的只涉及到 U 中属性的函数依赖集。

1) 自反性: 若 $Y \subseteq X \subseteq U$ 则 $X \rightarrow Y$ 在 R 上成立。

2) 增广性: 若 $X \rightarrow Y$ 在 R 上成立, 且 $Z \subseteq U$ 则 $XZ \rightarrow YZ$ 在 R 上成立。

3) 传递性: 若 $X \rightarrow Y$ 和 $Y \rightarrow Z$ 在 R 上成立, 则 $X \rightarrow Z$ 在 R 上成立。

6. FD 的其他五条推理规则:

1) 合并性: $\{X \rightarrow Y, X \rightarrow Z\} \models X \rightarrow YZ$

2) 分解性: $\{X \rightarrow Y, Z \subseteq Y\} \models X \rightarrow Z$

3) 伪传递性: $\{X \rightarrow Y, WY \rightarrow Z\} \models WX \rightarrow Z$

4) 复合性: $\{X \rightarrow Y, W \rightarrow Z\} \models WX \rightarrow YZ$

5) $\{X \rightarrow Y, W \rightarrow Z\} \models X \cup (W - Y) \rightarrow YZ$

7. 对于 FD $X \rightarrow Y$, 如果 $Y \subseteq X$, 那么称 $X \rightarrow Y$ 是一个“平凡的 FD”, 否则称为“非平凡的 FD”。通常研究非平凡 FD。

例: $X \rightarrow X$, $X \rightarrow \varnothing$, $\varnothing \rightarrow \varnothing$, $XY \rightarrow X$ 都是平凡函数依赖; $X \rightarrow XY$ 则是非平凡函数依赖。

8. 函数依赖是关键码概念的推广。

设关系模式 R 的属性集是 U, X 是 U 的一个子集。如果 $X \rightarrow U$ 在 R 上成立, 那么称 X 是 R 的一个超键。如果 $X \rightarrow U$ 在 R 上成立, 但对于 R 的任一真子集 X_1 都有 $X_1 \rightarrow U$ 不成立, 那么称 X 是 R 的一个候选键。在关系模式设计理论中, 键通常是指候选键。

9. 属性集的闭包

10. 设 F 是属性集 U 上的 FD 集, X 是 U 的子集, 那么 (相对于) 属性集 X 的闭包用 X^+ 表示, 它是一个从 F 集使用 FD 推理规则推出的所有满足 $X \rightarrow A$ 的属性 A 的集合: $X^+ = \{A \mid F \models X \rightarrow A\}$

11. $X \rightarrow Y$ 能用 FD 推理规则推出的充分必要条件是 $Y \subseteq X^+$, 从而避开求 F^+ , 使问题得到简化。

12. 求属性集 X 相对于 FD 集 F 的闭包 X^+ 的算法:

$X^+ = X;$

do $\{oldX^+ := X^+;$

for F 中每个 FD $Y \rightarrow Z$ do

if $Y \subseteq X^+$ then $X^+ := X^+ \cup Z;$

}while($X^+ \neq oldX^+$);

例: 属性集 U 为 ABCD FD 集为 $\{A \rightarrow B, B \rightarrow C, D \rightarrow B\}$ 。求 A^+ 、 $(AD)^+$ 和

$(BD)^+$

$A^+ = ABC$

$(AD)^+ = ABCD$

$(BD)^+ = BCD$

13. 如果关系模式 $R(U)$ 上的两个函数依赖集 F 和 G, 有 $F^+ = G^+$, 则称 F 和 G 是等价的函数依赖集。

三、关系模式的分解特性

1. 关系模式的分解:

设有关系模式 $R(U)$, 属性集为 U, 而 R_1, R_2, \dots, R_k 都是 U 的子集, 并且有 $R_1 \cup R_2 \cup \dots \cup R_k = U$ 。关系

模式 R_1, R_2, \dots, R_k 的集合用 ρ 表示, $\rho = \{R_1, R_2, \dots, R_k\}$ 。用 ρ 代替 R 的过程称为关系模式的分解。这里 ρ 称为 R 的一个分解, 也称为数据库模式。

一般把上述的 R 称为泛关系模式, R 对应的当前值称为泛关系。数据库模式 ρ 对应的当前值称为数据库实例, 它由数据库模式中的每一个关系模式的当前值组成。我们用 $\sigma = \langle r_1, r_2, \dots, r_k \rangle$ 表示。

因此, 在计算机中数据并不是存储在泛关系 r 中, 而是存储在数据库 σ 中。

2. σ 和 r 是否等价, 即是否表示同样的数据。这个问题用“无损分解”特性表示。

在模式 R 上有一个 FD 集 F , 在 ρ 的每一个模式 R_i 上有一个 FD 集 F_i , 那么 $\{F_1, F_2, \dots, F_k\}$ 与 F 是否等价。这个问题用“保持依赖”特性表示。

四、范式

1. 范式: 衡量关系模式好坏的标准。

2. 数据库设计中最常用的是 3NF 和 BCNF。

3. 第一范式 (1NF): 如果关系模式 R 的每个关系 r 的属性值都是不可分的原子值, 那么称 R 是第一范式的模式。满足 1NF 的关系称为规范化的关系, 否则称为非规范化的关系。1NF 是关系模式应具备的最起码的条件。

4. 局部依赖和完全依赖: 对于 FD $W \rightarrow A$, 如果存在 $X \sqsubset W$ 有 $X \rightarrow A$ 成立, 那么称 $W \rightarrow A$ 是局部依赖 (A 局部依赖于 W); 否则称 $W \rightarrow A$ 是完全依赖。

5. 主属性和非主属性: 如果 A 是关系模式 R 的候选键中的属性, 那么称 A 是 R 的主属性; 否则称 A 是 R 的非主属性。

6. 第二范式 (2NF): 如果关系模式是 1NF, 且每个非主属性完全函数依赖于候选键, 那么称 R 是第二范式 (2NF) 的模式。

7. 分解成 2NF 模式集的算法:

设关系模式 $R(U)$, 主键是 W , R 上还存在 FD $X \rightarrow Z$, 并且 Z 是非主属性和 X

$\sqsubset W$, 那么 $W \rightarrow Z$ 就是一个局部依赖。此时应把 R 分解成两个模式:

$R_1(XZ)$, 主键是 X ;

$R_2(Y)$, 其中 $Y=U-Z$, 主键仍是 W , 外键是 X (参照 R_1)。

如果 R_1 和 R_2 还不是 2NF, 则重复上述过程, 一直到数据库模式中的每一个关系模式都是 2NF 为止。

8. 如果 $X \rightarrow Y$, $Y \rightarrow A$, 且 $Y \rightarrow X$ 和 $A \sqsubset Y$, 那么称 $X \rightarrow A$ 是传递依赖 (A 传递依赖于 X)。

9. 第三范式 (3NF): 如果关系模式 R 是 2NF, 且每个非主属性都不传递依赖于 R 的候选键, 那么称 R 是第三范式 (3NF) 的模式。

10. 分解成 3NF 模式集的算法:

设关系模式 $R(U)$, 主键是 W , R 上还存在 FD $X \rightarrow Z$ 。并且 Z 是非主属性, $Z \sqsubset X$, X 不是候选键, 这样 $W \rightarrow Z$ 就是一个传递依赖。此时应把 R 分解成两个模式:

$R_1(XZ)$, 主键是 X ;

$R_2(Y)$, 其中 $Y=U-Z$, 主键仍是 W , 外键是 X (参照 R_1)。

如果 R_1 和 R_2 还不是 3NF, 则重复上述过程, 一直到数据库模式中的每一个关系模式都是 3NF 为止。

11. 如果 R 是 3NF 模式, 那么 R 也是 2NF 模式。如果 R 是 2NF 模式, 那么 R 也是 1NF 模式。

12. BC 范式 (BCNF): 如果关系模式 R 是 1NF, 且每个属性都不传递依赖于 R 的候选键, 那么称 R 是 BCNF 的模式。

13. 如果 R 是 BCNF 模式, 那么 R 也是 3NF 模式。

14. 分解成 BCNF 模式集的算法能保持无损分解, 但不一定能保持 FD 集。而分解成 3NF 模式集的算法既能保持无损分解, 又能保持 FD 集。

15. 关系模式由 1NF 分解为 2NF, 消除了非主属性对键的局部函数依赖; 由 2NF 分解为 3NF, 消除了非主属性对键的传递函数依赖; 而 BCNF 则消除了每一属性对键的传递函数依赖。

16. 关系模式设计理论主要用于数据库的逻辑设计过程中。

第三章复习题

一、单项选择题

1. 由于关系模式设计不当所引起的插入异常指的是 (B)

- A) 两个事务并发地对同一关系进行插入而造成数据库不一致
- B) 由于键值的一部分为空而不能将有用的信息作为一个元组插入到关系中
- C) 未经授权的用户对关系进行了插入
- D) 插入操作因为违反完整性约束条件而遭到拒绝

2. 下面有关模式分解的叙述中, 不正确的是 (D)

- A) 若一个模式分解保持函数依赖, 则该分解一定具有无损连接性
 B) 若要求分解保持函数依赖, 那么模式分解可以达到3NF 但不一定能达到BCNF
 C) 若要求分解既具有无损连接性, 又保持函数依赖, 则模式分解可以达到 3NF 但不一定能达到BCNF
 D) 若要求分解具有无损连接性, 那么模式分解一定可以达到BCNF
3. 下述哪一条不是由于关系模式设计不当而引起的 (B)
 A) 数据冗余 B) 丢失修改 C) 插入异常 D) 修改异常
4. 根据数据库规范化理论, 下面命题中正确的是 (D)
 A) 若 $R \in 2NF$ 则 $R \in 3NF$
 B) 若 $R \in 3NF$ 则 R 不属于 BCNF
 C) 若 $R \in 3NF$ 则 $R \in BCNF$
 D) 若 $R \in BCNF$ 则 $R \in 3NF$
5. 若关系模式 $R \in 3NF$ 则下面最正确的说法是 (C)
 A) 某个主属性不传递依赖于码
 B) 某个非主属性不部分依赖于码
 C) 所有的非主属性都不传递依赖于码
 D) 所有的非主属性都不部分依赖于码
6. 给定关系模式 $R(U, F)$, 其中, U 是所有属性的集合, F 是 FD 集。如果 X, Y 是 U 的子集, 且 $X \rightarrow Y \in F$, 则 X 和 Y 之间必然存在 (C)
 A) 一对一联系
 B) 一对多联系 (含一对一联系)
 C) 多对一联系 (含一对一联系)
 D) 多对多联系
7. 设 $R(U)$, 其中, U 是所有属性的集合。如果存在 U 的子集 K , 且 $K \rightarrow U$, 则 K 为 R 的 (D)
 A) 外键 B) 候选键 C) 主键 D) 超键
8. 任何一个二元关系在函数依赖的范畴内必能达到 (D)
 A) 1NF B) 2NF C) 3NF D) BCNF
9. 在关系模式设计理论中, 如果一个关系 R 满足 1NF, 但 R 的某个非主属性传递依赖于键, 则关系 R 至多属于 (B)
 A) 1NF B) 2NF C) 3NF D) BCNF
10. 在一个 BCNF 关系模式中, 所有的非主属性对每一个键都是 (D)
 A) 部分函数依赖 B) 平凡函数依赖
 C) 传递函数依赖 D) 完全函数依赖
11. 在一个关系模式 $R(A, B, C, D)$ 中, 若各个属性间没有任何函数依赖关系, 则该模式的主属性有 (A)
 A) A, B, C, D B) R, A, C C) A, B, D D) R, A, B, C, D
12. 当下述哪一条成立时, 称 $X \rightarrow Y$ 为平凡的函数依赖 (B)
 A) $Y \subseteq X$ B) $Y \supseteq X$ C) $X \cap Y = \emptyset$ D) $X \cap Y \neq \emptyset$
13. 当关系模式 $R(A, B)$ 已属于 3NF 下列 (B) 说法是正确的。
 A) 它一定消除了插入和删除异常
 B) 仍可能存在着一一定的插入和删除异常
 C) 一定属于 BCNF
 D) A 和 C 都是
14. 关系模型中的关系模式至少是 (A)
 A) 1NF B) 2NF C) 3NF D) BCNF
15. 下列函数依赖中, (C) 是平凡的函数依赖。
 A) $AB \rightarrow BC$ B) $AB \rightarrow CD$ C) $AB \rightarrow A$ D) $AB \rightarrow D$
16. 下列命题中, 不正确的是 (D)
 A) 若 $X \rightarrow Y$ 在 R 上成立, 且 $Z \subseteq U$, 则 $XZ \rightarrow YZ$ 在 R 上成立。
 B) 若 $X \rightarrow Y$ 和 $Y \rightarrow Z$ 在 R 上成立, 则 $X \rightarrow Z$ 在 R 上成立。
 C) 若 $X \rightarrow Y$, $X \rightarrow Z$ 在 R 上成立, 则 $X \rightarrow YZ$ 在 R 上成立。
 D) 若 $X \rightarrow Y$, $WY \rightarrow Z$ 在 R 上成立, 则 $WX \rightarrow Z$ 在 R 上不成立。
17. 设关系模式 $R(ABCDE)$, F 是 R 上成立的 FD 集, $F = \{AB \rightarrow C, CD \rightarrow E, DE \rightarrow B\}$, 则下列哪一项不是

关系模式 R 的候选键 (D)

A) ACD B) ABD C) AED D) AD

18. 设关系模式 R (ABCD) 上 FD 集为 F, 并且 $F = \{ AB \rightarrow C, C \rightarrow D, D \rightarrow A \}$, 则下列哪一项不是关系模式 R 的候选键 (B)

A) AB B) AD C) BC D) BD

二、填空题

1. 关系模式规范化过程中, 若要求分解保持函数依赖, 那么模式分解一定可以达到 3NF 但不一定能达到 BCNF

2. 将一个关系从 1NF 规范到 2NF 目的是消除非主属性对键的部分函数依赖, 若进一步规范到 3NF 目的是消除非主属性对键的传递函数依赖。

3. 在关系数据库的规范化设计中, 对模式分解的等价性进行评价的两条主要标准是具有无损连接性和保持函数依赖。

4. 若关系为 1NF 且它的每一非主属性都完全函数依赖于候选键, 则该关系为 2NF。

5. 衡量关系模式好坏的标准称为范式。

6. 满足第二范式的关系称为规范化的关系。