

【软考达人】

软考资料免费获取

- 1、最新软考题库
- 2、软考备考资料
- 3、考前压轴题



微信扫一扫，立马获取



6W+免费题库



免费备考资料

PC版题库: ruankaodaren.com

试题 1(2017 年上半年试题 40)

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

试题分析

本题考查的是数据库概念结构设计具体步骤。题目选项所展示的四个步骤中，有两个是我们熟知的：设计局部视图和合并取消冲突。所以解题的关键点，是分析清楚另外两个步骤到底是完成什么任务。从题目选项来看，无非是分析抽象数据与设计局部视图谁先谁后，以及合并取消冲突与修改重构消除冗余谁先谁后的问题。

抽象数据是将实际数据的特征提取出来以便建立模型，所以抽象数据应在设计局部视图之前。

而修改重构消除冗余应在合并取消冲突之后，因为重构往往意味着在调优，调优是需要先有雏形的。

试题答案

(40) C

试题 2(2017 年上半年试题 41)

设有员工关系 Emp (员工号，姓名，性别，年龄，电话，家庭住址，家庭成员，关系，联系电话)。其中，“家庭成员，关系，联系电话”分别记录了员工亲属的姓名、与员工的关系以及联系电话，且一个员工允许有多个家庭成员。为使

数据库模式设计更合理，对于员工关系 Emp ()。

- A.只允许记录一个亲属的姓名、与员工的关系以及联系电话
- B.可以不作任何处理，因为该关系模式达到了 3NF
- C.增加多个家庭成员、关系及联系电话字段
- D.应该将家庭成员、关系及联系电话加上员工号设计成一个独立的模式

试题分析

题目中员工允许有多个家庭成员，此时如果要把多个家庭成员都记录下来，必然造成表中数据的冗余。此时采用拆分表格的方式来处理是最理想的状态。

试题答案

(41) D

试题 3(2017 年上半年试题 42-43)

给定关系模式 $R < U, F >$ ， $U = \{A, B, C, D, E\}$ ， $F = \{B \rightarrow A, D \rightarrow A, A \rightarrow E, AC \rightarrow B\}$ ，则 R 的候选关键字为 ()，分解 $\rho = \{R_1(ABCE), R_2(CD)\}$ ()。

- A.CD
- B.ABD
- C.ACD
- D.ADE

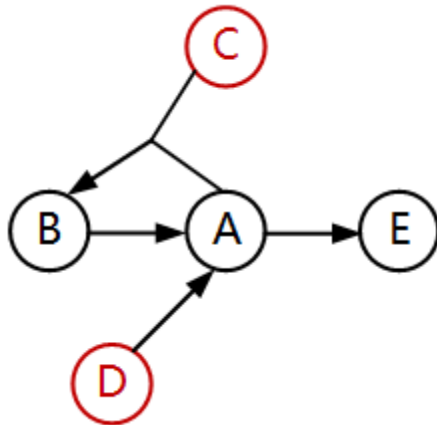
- A.具有无损连接性，且保持函数依赖
- B.不具有无损连接性，但保持函数依赖

C.具有无损连接性，但不保持函数依赖

D.不具有无损连接性，也不保持函数依赖

试题分析

根据题意可绘制出函数依赖图：



从图可以看出，入度为 0 的结点只有 C 与 D，而两者的结合能遍历全图，所以 CD 为候选键。

用公式法判断是否为无损分解，由于 $R_1 \cap R_2 = C$ ，而原关系中没有 $C \rightarrow^*$ 的函数依赖，所以可以断定，分解为有损分解。

判断是否保持函数依赖，由于函数依赖 $D \rightarrow A$ 未被保存下来，所以此分解不是保持函数依赖的分解。

试题答案

(42) A (43) D

试题 4(2017 年上半年试题 44-45)

在关系 $R(A_1, A_2, A_3)$ 和 $S(A_2, A_3, A_4)$ 上进行

$\pi_{A_1, A_4}(\sigma_{A_2 < '2017' \wedge A_4 = '95'}(R \bowtie S))$ 关系运算，与该关系表达式等价的是

()。

将该关系代数表达式转换为等价的 SQL 语句如下：

SELECT A1, A4 FROM R, S WHERE R.A2 < '2017' () A.

$$\pi_{1,4}(\sigma_{2<'2017' \vee 4='95'}(R \bowtie S))$$

B. $\pi_{1,6}(\sigma_{2<'2017'}(R) \times \sigma_{3='95'}(S))$

C. $\pi_{1,4}(\sigma_{2<'2017'}(R) \times \sigma_{6='95'}(S))$

D. $\pi_{1,6}(\sigma_{2=4 \wedge 3=5}(\sigma_{2<'2017'}(R) \times \sigma_{3='95'}(S)))$

A. OR S.A4 < '95' OR R.A2 = S.A2 OR R.A3 = S.A3

B. AND S.A4 < '95' OR R.A2 = S.A2 AND R.A3 = S.A3

C. AND S.A4 < '95' AND R.A2 = S.A2 AND R.A3 = S.A3

D. OR S.A4 < '95' AND R.A2 = S.A2 OR R.A3 = S.A3

试题分析

本题前一空分析：

A 选项的关系代数表达式，错误之处在于选择的两个条件不应为“或”关系。

B 选项的关系代数表达式，错误之处在于 R 与 S 仅做了笛卡尔积的操作，并没有把相同属性列做等值判断。应加上：2=4 及 3=5 的选择条件才对。

C 选项的关系代数表达式，与 B 选项有相同错误，同时投影列号还不正确。

后面一空，首先值得说明的是，选项中有一个共同的错误，即“S.A4 < '95' ”应修改为“S.A4 = '95' ”。

选项中几个条件都是正确的，需要选择的，其实只是使用 AND 还是 OR

来进行连接。由于进行自然连接以及相关条件判断都是同时要成立的，所以必须都要用 AND 进行连接。

试题答案

(44) D (45) C

试题 5(2016 年上半年试题 22-23)

商业智能关注如何从业务数据中提取有用的信息，然后采用这些信息指导企业的业务开展。商业智能系统主要包括数据预处理、（ ）、数据分析和数据展现 4 个主要阶段。其中数据预处理是整合企业原始数据的第一步，它包括数据的抽取、

（ ）和加载三个过程。A.数据索引

B.数据仓库

C.数据库

D.数据字典

A.转换

B.分析

C.查询

D.建模

试题分析

BI 系统主要包括数据预处理、建立数据仓库、数据分析和数据展现四个主要阶段。

数据预处理是整合企业原始数据的第一步，它包括数据的抽取（Extraction）、转换（Transformation）和加载（Load）三个过程（ETL 过程）；建立数据仓

库则是处理海量数据的基础；数据分析是体现系统智能的关键，一般采用 OLAP 和数据挖掘两大技术。OLAP 不仅进行数据汇总/聚集，同时还提供切片、切块、下钻、上卷和旋转等数据分析功能，用户可以方便地对海量数据进行多维分析。数据挖掘的目标则是挖掘数据背后隐藏的知识，通过关联分析、聚类和分类等方法建立分析模型，预测企业未来发展趋势和将要面临的问题；在海量数据和分析手段增多的情况下，数据展现则主要保障系统分析结果的可视化。

试题答案

(22) B (23) A

试题 6(2016 年上半年试题 40)

在采用三级模式结构的数据库系统中，如果对数据库中的表 Emp 创建聚簇索引，那么改变的是数据库的（ ）。A.模式

B.内模式

C.外模式

D.用户模式

试题分析

索引的写入修改了数据库的物理结构，而不是简单的逻辑设计。内模式规定了数据在存储介质上的物理组织方式、记录寻址方式。

试题答案

(40) B

试题 7(2016 年上半年试题 41)

分布式事务的执行可能会涉及到多个站点上的数据操作，在两阶段提交协议中，当事务 T_i 的所有读写操作执行结束后，事务 T_i 的发起者协调器 C_i 向所有参与 T_i

的执行站点发送<prcpare T_i >的消息，当收到所有执行站点返回的<ready T_i >消息后， C_i 再向所有执行站点发送<commit T_i >消息。若参与事务 T_i 执行的某个站点故障恢复后日志中有<ready T_i >记录，而没有<commit T_i >记录，则

- ()。A.事务 T_i 已完成提交，该站点无需做任何操作
B.事务 T_i 已完成提交，该站点应做 REDO 操作
C.事务 T_i 未完成提交，该站点应做 UNDO 操作
D.应向协调器询问以决定 T_i 的最终结果

试题分析

两阶段提交协议受如下两条规则的支配：1、只要有一个参与者撤销事务，协调者就必须做出全局撤销决定。2、只有所有参与者都同意提交事务，协调者才能做出全局提交决定。题干中，事务日志中只有<ready T_i >记录，说明已经完成了准备提交阶段，而没有<commit T_i >记录，说明该事务没有进入提交阶段，所以，需要向协调器去查询 T_i 的最终决定结果。

试题答案

(41) D

试题 8(2016 年上半年试题 42-43)

假设关系 $R(A_1, A_2, A_3)$ 上的一个分解为 $\rho = \{(A_1, A_2), (A_1, A_3)\}$ ，下表是关系 R 上的一个实例，根据实例推出 R 的函数依赖集 F 为 ()，分解 ρ ()。

R		
A_1	A_2	A_3
a	a	d
a	b	e
a	c	f

$$A.F=\{A_1\rightarrow A_2\}$$

$$B.F=\{A_1A_3\rightarrow A_2, A_1A_2\rightarrow A_3\}$$

$$C.F=\{A_1\rightarrow A_3\}$$

$$D.F=\{A_1\rightarrow A_2, A_1\rightarrow A_3\}$$

A.是无损联接的

B.是保持函数依赖的

C.是有损联接的

D.无法确定是否保持函数依赖

试题分析

由于 A_1 为 a 时, A_2 可能是 a 或 b 或 c , 所以可以确定 $A_1 \rightarrow A_2$ 不成立。同理 $A_1 \rightarrow A_3$ 不成立。此时 ACD 三个选项均可以排除。

$$(A_1A_2) \cap (A_1A_3) = A_1$$

$$(A_1A_2) - (A_1A_3) = A_2$$

$$(A_1A_3) - (A_1A_2) = A_3$$

由于 $A_1 \rightarrow A_2$ 与 $A_1 \rightarrow A_3$ 均不成立, 所以有损。

试题答案

(42) B (43) C

试题 9(2016 年上半年试题 44-45)

给定关系 $R(A,B,C,D)$ 和关系 $S(C,D,E)$, 对其进行自然连接运算 $R \bowtie S$ 后的属性列

为 () 个; 与 $\sigma_{R.B > S.E}(R \bowtie S)$ 等价的关系代数表达式为 ()。A.4

B.5

C.6

D.7

A. $\sigma_{2>7}(R \times S)$

B. $\pi_{1,2,3,4,7}(\sigma_{'2' > '7' \wedge 3=5 \wedge 4=6}(R \times S))$

C. $\sigma_{'2' > '7'}(R \times S)$

D. $\pi_{1,2,3,4,7}(\sigma_{2>7 \wedge 3=5 \wedge 4=6}(R \times S))$

试题分析

对 R 与 S 做自然联接运算时，联接条件为：R.C=S.C and R.D=S.D。运算结果会自动去重复列，所以结果的列为：A、B、C、D、E，一共 5 列。

试题答案

(44) B (45) D

试题 10(2015 年上半年试题 23)

采用数据仓库技术进行数据收集时，有时会遇一些略微不一致但可以纠正的数据，纠正这些数据的过程称为（ ）。A.数据转换

B.数据抽取

C.数据清洗

D.数据装载

试题分析

数据清洗从名字上也看的出就是把“脏”的“洗掉”，指发现并纠正数据文件中可识别的错误的最后一道程序，包括检查数据一致性，处理无效值和缺失值等。

因为数据仓库中的数据是面向某一主题的数据的集合，这些数据从多个业务系统中抽取而来，而且包含历史数据，这样就避免不了有的数据是错误数据、有的数据相互之间有冲突，这些错误的或有冲突的数据显然是我们不想要的，称为“脏数据”。我们要按照一定的规则把“脏数据”“洗掉”，这就是数据清洗。而数据清洗的任务是过滤那些不符合要求的数据，将过滤的结果交给业务主管部门，确认是否过滤掉还是由业务单位修正之后再进行抽取。不符合要求的数据主要有不完整的数据、错误的数据、重复的数据三大类。数据清洗是与问卷审核不同，录入后的数据清理一般是由计算机而不是人工完成。

试题答案

(23) C

试题 11(2015 年上半年试题 40)

数据仓库中数据（ ）的特点是指数据一旦进入数据仓库后，将被长期保留并定期加载和刷新，可以进行各种查询操作，但很少对数据进行修改和删除操作。

- A.面向主题
- B.集成性
- C.相对稳定性
- D.反映历史变化

试题分析

数据仓库有四大特点：

1、面向主题

操作型数据库的数据组织面向事务处理任务，各个业务系统之间各自分离，而数据仓库中的数据是按照一定的主题域进行组织的。

2、集成的

数据仓库中的数据是在对原有分散的数据库数据抽取、清理的基础上经过系统加工、汇总和整理得到的，必须消除源数据中的不一致性，以保证数据仓库内的信息是关于整个企业的一致全局信息。

3、相对稳定的

数据仓库的数据主要供企业决策分析之用，所涉及的数据操作主要是数据查询，一旦某个数据进入数据仓库以后，一般情况下将被长期保留，也就是数据仓库中一般有大量的查询操作，但修改和删除操作很少，通常只需要定期的加载、刷新。

4、反映历史变化

数据仓库中的数据通常包含历史信息，系统记录了企业从过去某一时点(如开始应用数据仓库的时点)到目前的各个阶段的信息，通过这些信息，可以对企业的发展历程和未来趋势做出定量分析和预测。

试题答案

(40) C

试题 12(2015 年上半年试题 41)

数据挖掘的分析方法可以划分为关联分析、序列模式分析、分类分析和聚类分析四种。如果某方法需要一个示例库（该库中的每个元组都有一个给定的类标识）做训练集时，这种分析方法属于（ ）。A.关联分析

B.序列模式分析

C.分类分析

D.聚类分析

试题分析

关联分析、序列分析、分类分析、聚类分析的基本思想为：。

(1) 关联分析。关联分析主要用于发现不同事件之间的关联性，即一个事件发生的同时，另一个事件也经常发生。关联分析的重点在于快速发现那些有实用价值的关联发生的事件。其主要依据是事件发生的概率和条件概率应该符合一定的统计意义。在进行关联分析的同时，还需要计算两个参数，分别是最小置信度（可信度）和最小支持度，前者表示规则需满足的最低可靠度，用以过滤掉可能性过小的规则；后者则用来表示规则在统计意义上需满足的最小程度。

(2) 序列分析。序列分析主要用于发现一定时间间隔内接连发生的事件，这些事件构成一个序列，发现的序列应该具有普遍意义，其依据除了统计上的概率之外，还要加上时间的约束。在进行序列分析时，也应计算置信度和支持度。

(3) 分类分析。分类分析通过分析具有类别的样本特点，得到决定样本属于各种类别的规则或方法。利用这些规则和方法对未知类别的样本分类时应该具有一定的准确度。其主要方法有基于统计学的贝叶斯方法、神经网络方法、决策树方法等。分类分析时首先为每个记录赋予一个标记（一组具有不同特征的类别），即按标记分类记录，然后检查这些标定的记录，描述出这些记录的特征。这些描述可能是显式的，例如，一组规则定义；也可能是隐式的，例如，一个数学模型或公式。

(4) 聚类分析。聚类分析是根据“物以类聚”的原理，将本身没有类别的样本聚集成不同的组，并且对每个这样的组进行描述的过程。其主要依据是聚集到同一个组中的样本应该彼此相似，而属于不同组的样本应该足够不相似。聚类分析法是分类分析法的逆过程，它的输入集是一组未标定的记录，即输入的记录没有

作任何处理，目的是根据一定的规则，合理地划分记录集合，并用显式或隐式的方法描述不同的类别。

试题答案

(41) C

试题 13(2015 年上半年试题 42-43)

给定关系模式 $R(A_1, A_2, A_3, A_4)$ ， R 上的函数依赖集 $F=\{A_1A_3 \rightarrow A_2,$

$A_2 \rightarrow A_3\}$ ，则 R ()。若将 R 分解为 $p=\{(A_1, A_2), (A_1, A_3)\}$ ，那么该分解

()。A.有一个候选关键字 A_1A_3

B.有一个候选关键字 $A_1A_2A_3$

C.有两个候选关键字 $A_1A_3A_4$ 和 $A_1A_2A_4$

D.有三个候选关键字 A_1A_2 、 A_1A_3 和 A_1A_4

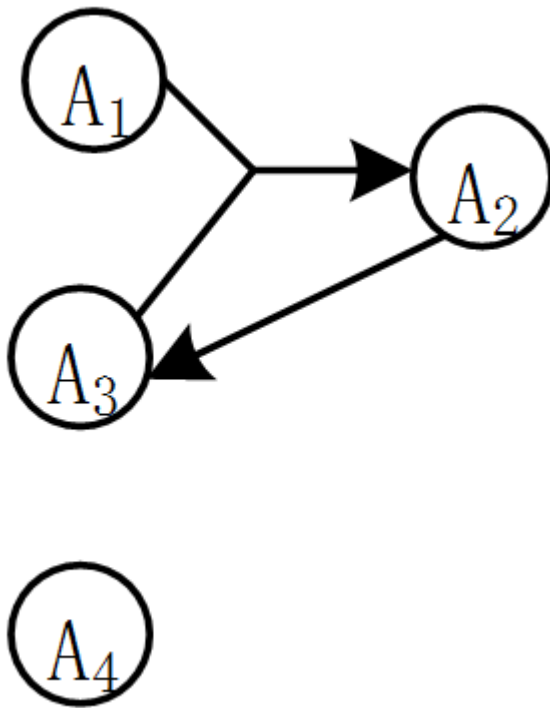
A.是无损连接的

B.是保持函数依赖的

C.既是无损连接又保持函数依赖

D.既是有损连接又不保持函数依赖

试题分析



在判断无损分解时，我们采用公式法：

R1 与 R2 的交集是：A1。R1-R2=A2，R2-R1=A3，由于 A1→A2 和 A1→A3，均不成立，所以有损。

是否保持函数依赖，就看函数依赖两边的属性是否在分解后的关系中都有。

A1A3→A2，A2→A3 都没有被保存下来，所以没有保持函数依赖。

试题答案

(42) C (43) D

试题 14(2015 年上半年试题 44-45)

关系 R、S 如下表所示， $R \div (\pi_{A1,A2}(\sigma_{1<3}(S)))$ 的结果为 ()，R、S 的左外联接、右外联接和完全外联接的元组个数分别为 ()。

R		
A1	A2	A3
1	2	3
2	1	4
3	4	4
4	6	7

S		
A1	A2	A4
1	9	1
2	1	8
3	4	4
4	8	3

- A.{4}
- B.{3,4}
- C.{3,4,7}
- D.{(1,2),(2,1),(3,4),(4,7)}

- A.2,2,4
- B.2,2,6
- C.4,4,4
- D.4,4,6

试题分析

首先求解的是 $R \div (\pi_{A1,A2}(\sigma_{1 < 3}(S)))$ 的结果, 我们先将 $(\pi_{A1,A2}(\sigma_{1 < 3}(S)))$ 的结果求出, 为:

A1	A2
2	1
3	4

然后确定除法结果集的属性包括哪些。这个属性是：被除关系的属性集 - 除数关系的属性集得来的, 在此 $R(A1,A2,A3)-(A1,A2)=A3$, 所以结果集应是单属性, 从这里就已经可以排除 D 选项了。

接下来结果集要求与除数记录 “2, 1” 和 “3, 4” 拼接成的记录在原关系 R 中都能找到。符合这个条件的，只有{4}。

外联接可以是左向外联接、右向外联接或完全外部联接。

在 FROM 子句中指定外联接时，可以由下列几组关键字中的一组指定：

1) LEFT JOIN 或 LEFT OUTER JOIN

左向外联接的结果集包括 LEFT OUTER 子句中指定的左表的所有行，而不仅仅是联接列所匹配的行。如果左表的某行在右表中没有匹配行，则在相关联的结果集行中右表的所有选择列表列均为空值。

2) RIGHT JOIN 或 RIGHT OUTER JOIN

右向外联接是左向外联接的反向联接。将返回右表的所有行。如果右表的某行在左表中没有匹配行，则将为左表返回空值。

3) FULL JOIN 或 FULL OUTER JOIN

完全外部联接返回左表和右表中的所有行。当某行在另一个表中没有匹配行时，则另一个表的选择列表列包含空值。如果表之间有匹配行，则整个结果集行包含基表的数据值。

在本题中，左表为 R，右表为 S，他们的记录数均为 4，所以左外连接和右外连接的记录数均为 4，全连接时，会将左表与右表中的记录均列出来，但左表与右表均有 A1 与 A2 为 “2, 1” 和 “3, 4” 所以一共会有 6 条记录。

试题答案

(44) A (45) D

试题 15(2014 年上半年试题 5)

使用 UML 进行关系数据库的（ ）时，需要设计出表达持久数据的实体类及其联系，并将它们映射为数据库表和视图等。

- A.业务用例设计
- B.逻辑数据模型设计
- C.物理数据模型设计
- D.物理实现设计

试题分析

本题考查关系型数据库的设计问题。数据库设计包括：概念设计、逻辑设计、物理设计。其中数据表这个级别的设计归属于逻辑设计阶段。

试题答案

(5) B

试题 16(2014 年上半年试题 40-41)

在数据库系统中，数据库的视图、基本表和存储文件的结构分别与（ ）对应；数据的物理独立性和数据的逻辑独立性是分别通过修改（ ）来完成的。

- A.模式、外模式、内模式
- B.模式、内模式、外模式
- C.外模式、模式、内模式
- D.外模式、内模式、模式

- A.模式与内模式之间的映像、外模式与模式之间的映像
- B.外模式与内模式之间的映像、外模式与模式之间的映像

C.外模式与模式之间的映像、模式与内模式之间的映像

D.外模式与内模式之间的映像、模式与内模式之间的映像

试题分析

本题考查对数据库基本概念掌握程度。

数据库通常采用三级模式结构，其中，视图对应外模式，基本表对应模式，存储文件对应内模式。

数据的独立性是由 DBMS 的二级映像功能来保证的。数据的独立性包括数据的物理独立性和数据的逻辑独立性。数据的物理独立性是指当数据库的内模式发生改变时，数据的逻辑结构不变。为了保证应用程序能够正确执行，需要通过修改概念模式与内模式之间的映像。数据的逻辑独立性是指用户的应用程序与数据库的逻辑结构是相互独立的。数据的逻辑结构发生变化后，用户程序也可以不修改。但是，为了保证应用程序能够正确执行，需要修改外模式与概念模式之间的映像。

试题答案

(40) C (41) A

试题 17(2014 年上半年试题 42-43)

给定关系模式 $R(U, F)$ ， $U = \{A, B, C, D\}$ ， $F = \{AB \rightarrow C, CD \rightarrow B\}$ 。关系 R ()，且分别有 ()。

A.只有 1 个候选关键字 ACB

B.只有 1 个候选关键字 BCD

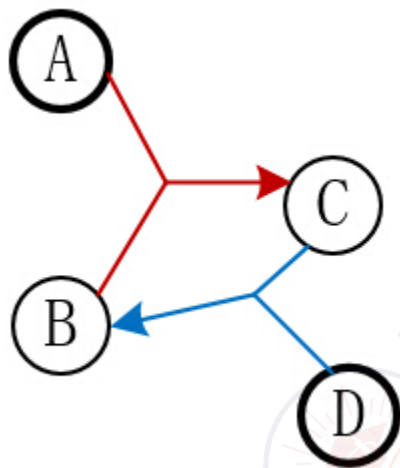
C.有 2 个候选关键字 ACD 和 ABD

D.有 2 个候选关键字 ACB 和 BCD

- A.0 个非主属性和 4 个主属性
- B.1 个非主属性和 3 个主属性
- C.2 个非主属性和 2 个主属性
- D.3 个非主属性和 1 个主属性

试题分析

解决这类问题，使用图示法比较合适，以下是用图示法的方式画出的图：



A 与 D 是入度为零的结点，他们一定是候选键的组成部分。但仅有 A 与 D 还不能充当候选键，所以要加入其它的属性；加入 B 之后，ABD 能遍历全图，能充当候选键；加入 C 之后，ACD 也能遍历全图，所以 ABD 与 ACD 都是候选键。求出候选键之后，包含于任意候选键中的所有属性都是主属性，所以 ABCD 都是主属性。

试题答案

(42) C (43) A

试题 18(2014 年上半年试题 44-45)

某销售公司数据库中的关系零件为 $P(Pno, Pname, Sname, City, Qty)$, Pno 表示零件号, $Pname$ 表示零件名称, $Sname$ 表示供应商, $City$ 表示所在地, Qty 表示库存量。其函数依赖集 $F=\{Pno \rightarrow Pname, (Pno, Sname) \rightarrow Qty, Sname \rightarrow City\}$ 。关系 P 为 (), 存在冗余度大、修改操作不一致、插入异常和删除异常的问题。若将 P 分解为 (), 则可以解决这一问题。

A.1NF

B.2NF

C.3NF

D.4NF

A. $P_1(Pname, Qty)$ 、 $P_2(Pno, Sname, City)$

B. $P_1(Pname, Pname)$ 、 $P_2(Sname, City, Qty)$

C. $P_1(Pno, Pname)$ 、 $P_2(Pno, Sname, Qty)$ 、 $P_3(Sname, City)$

D. $P_1(Pno, Pname)$ 、 $P_2(Pno, Qty)$ 、 $P_3(Sname, City)$ 、 $P_4(City, Qty)$

试题分析

本题考查数据库基本概念、数据库设计基础知识。

原零件关系 P 存在非主属性对码的部分函数依赖: $(Pno, Sname) \twoheadrightarrow Qty$, 但是 $Pno \rightarrow Pname$ 、 $Sname \rightarrow City$, 因此 P 属于 1NF, 而非 2NF 的。1NF 主要存在冗余度大、修改操作的不一致、插入异常和删除异常的问题。

分解后的关系模式 P_1 、 P_2 和 P_3 消除了非主属性对码的部分函数依赖, 同时不存在传递依赖, 故达到 3NF。

试题答案

(44) A (45) C

试题 19(2013 年上半年试题 23-24)

商业智能系统主要包括数据预处理、建立数据仓库、数据分析和数据展现 4 个主要阶段，其中（ ）是处理海量数据的基础；数据分析是体现系统智能的关键，一般采用（ ）和数据挖掘技术。

A.数据预处理

B.建立数据仓库

C.数据分析

D.数据展现

A.联机分析处理

B.联机事务处理

C.脱机分析处理

D.脱机事务处理

试题分析

BI 系统主要包括数据预处理、建立数据仓库、数据分析和数据展现四个主要阶段。

数据预处理是整合企业原始数据的第一步，它包括数据的抽取（Extraction）、转换（Transformation）和加载（Load）三个过程（ETL 过程）；建立数据仓库则是处理海量数据的基础；数据分析是体现系统智能的关键，一般采用联机分析处理（OLAP）和数据挖掘两大技术。联机分析处理不仅进行数据汇总/聚集，同时还提供切片、切块、下钻、上卷和旋转等数据分析功能，用户可以方便地对

海量数据进行多维分析。数据挖掘的目标则是挖掘数据背后隐藏的知识，通过关联分析、聚类 and 分类等方法建立分析模型，预测企业未来发展趋势和将要面临的问题；在海量数据和分析手段增多的情况下，数据展现则主要保障系统分析结果的可视化。

试题答案

(23) B (24) A

试题 20(2013 年上半年试题 40)

在数据库系统中，为了保证数据库的（ ），通常由 DBA 使用 DBMS 提供的授权功能为不同用户授权。

- A.可靠性
- B.安全性
- C.一致性
- D.完整性

试题分析

在数据库中 DBA 是拥有最高权限的管理员，他可以根据不同用户的数据需求，给予不同的访问授权，这样能有效的控制用户只能访问自己权限以内的数据，从而保障整个数据库的安全性。

试题答案

(40) B

试题 21(2013 年上半年试题 41-42)

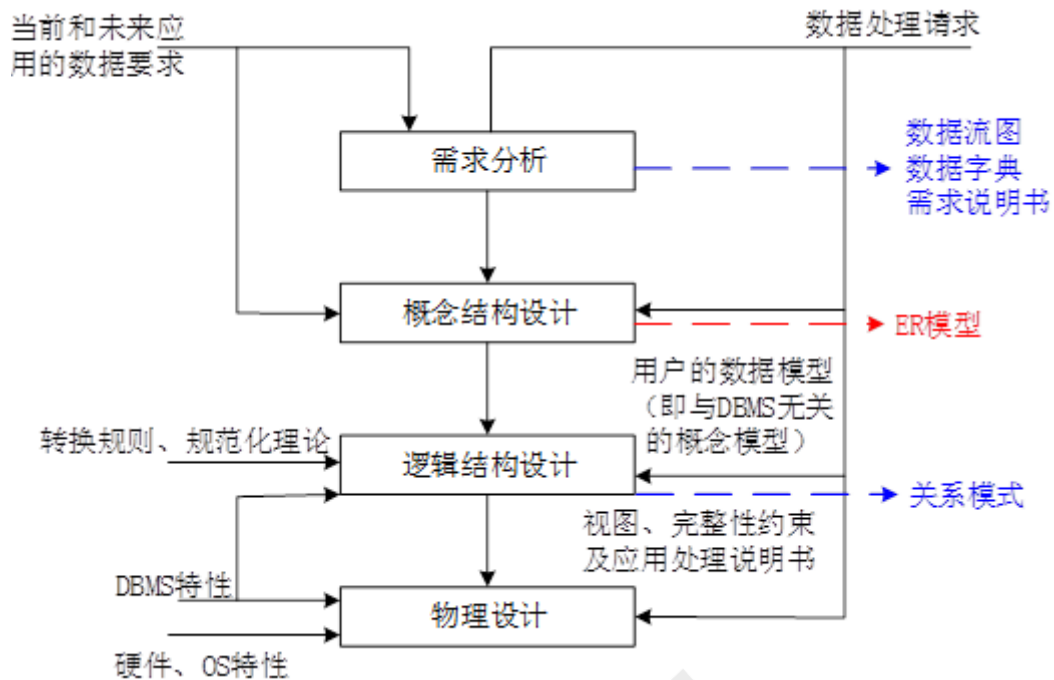
在数据库设计的需求分析、概念结构设计、逻辑结构设计和物理结构设计的四个阶段中，基本 E-R 图是（ ）；数据库逻辑结构设计阶段的主要工作步骤依次为（ ）。

- A.需求分析阶段形成的文档，并作为概念结构设计阶段的设计依据
- B.逻辑结构设计阶段形成的文档，并作为概念结构设计阶段的设计依据
- C.概念结构设计阶段形成的文档，并作为逻辑结构设计阶段的设计依据
- D.概念结构设计阶段形成的文档，并作为物理设计阶段的设计依据

- A.关系规范化→转换为数据模型→模式优化→设计用户模式
- B.转换为数据模型→关系规范化→模式优化→设计用户模式
- C.模式优化→设计用户模式→关系规范化→转换为数据模型
- D.设计用户模式→模式优化→关系规范化→转换为数据模型

试题分析

本题考查数据库的分析与设计过程，每个阶段结束的时候都有相应的产出物，如图所示。



从图可以看出，ER图是概念结构设计阶段的产出物。

数据库逻辑结构设计阶段的任务是将ER图转成数据模式（如：关系模式），然后通过规范化将其结构调整好，并做相应的优化，最后是形成用户模式。

试题答案

(41) C (42) B

试题 22(2013 年上半年试题 43-45)

给定关系模式科室 K（科室号，科室名，负责人，科室电话）、医生 Y（医生号，医生名，性别，科室号，联系电话，家庭地址）和患者 B（病历号，患者名，性别，医保号，联系方式），并且 1 个科室有多名医生，1 名医生属于 1 个科室；1 名医生可以为多个患者诊疗，1 个患者也可以找多名医生诊疗。

科室与医生之间的“所属”联系类型、医生与患者之间的“诊疗”联系类型分别为（ ）；其中（ ）。下列查询“肝胆科”医生的医生名、联系电话及家庭住址的关系代数表达式中，查询效率最高的是（ ）。

A.1:1、n:m

B.n:m、1:1

C.n:m、1:n

D.1:n、n:m

A. “诊疗” 联系需要转换为一个独立的关系，并将医生号和患者名作为主键

B. “诊疗” 联系需要转换为一个独立的关系，并将医生号和病历号作为主键

C. “所属” 联系需要转换为一个独立的关系，并将医生号和科室名作为主键

D. “所属” 联系需要转换为一个独立的关系，并将医生号和科室号作为主键

A. $\pi_{6,9,10}(\sigma_{1=8}(K \times Y))$

B. $\pi_{6,9,10}(\sigma_{1=8 \wedge 2='肝胆科'}(K \times Y))$

C. $\pi_{6,8,9}(\sigma_{2='肝胆科'}(K \bowtie Y))$

D. $\pi_{3,4,5}(\pi_1(\sigma_{2='肝胆科'}(K)) \bowtie Y)$

试题分析

本题考查规范化与关系代数相关知识。

由于 1 个科室有多名医生，1 名医生属于 1 个科室，所以科室与医生的对应联系类型为：1:n；1 名医生可以为多个患者诊疗，1 个患者也可以找多名医生诊疗，所以医生与患者的对应联系类型为：n:m。

在将联系转成关系模式时, 1:1 和 1:n 的联系都可以不单独转成关系模式, 而 m:n 联系必须单独转成一个关系模式。这就决定了: 医生与患者之间的“诊疗”联系需要单独转成一个关系模式, 这个关系模式以医生 Y 与患者 B 两个关系模式的主键组合作为主键, 即医生号和病历号作为主键。

最后一空为关系代数表达式查询问题, 备选答案中, 能达到题目效果的是 B 与 C。相对而言 C 的效率比 B 高。

C 选项中, 先将 K 与 Y 做自然连接, 得到的结果属性包括: 科室号, 科室名, 负责人, 科室电话, 医生号, 医生名, 性别, 联系电话, 家庭地址。然后做选择操作, 将科室名为“肝胆科”的医生记录选出, 最后做投影操作, 将结果集的属性限制为: 医生名、联系电话及家庭住址。

试题答案

(43) D (44) B (45) C

试题 23(2012 年上半年试题 39)

数据库的视图与基本表之间, 基本表与存储文件之间分别通过建立 () 之间的映像, 保证数据的逻辑独立性和物理独立性。

- A. 模式到内模式和外模式到内模式
- B. 外模式到内模式和内模式到模式
- C. 外模式到模式和模式到内模式
- D. 内模式到模式和模式到外模式

试题分析

本题考查数据库系统基本概念。在数据库系统中有三级模式: 外模式、模式和内模式。外模式也称用户模式或子模式, 用于描述用户视图层次上的数据特性;

模式用于对数据库中全部数据的逻辑结构和特征进行描述,即模式用于描述概念视图层次上的数据特性,如数据库中的基本表;内模式用于描述内部视图层次上的数据特性,是数据在数据库内部的表示方式,如存储文件。数据库的视图与基本表之间通过外模式到模式之间的映像,实现了外模式到概念模式之间的相互转换,即实现了视图与基本表之间的相互转换,从而保证了数据的逻辑独立性。数据库的基本表与存储文件之间通过模式到内模式之间的映像,实现了概念模式到内模式之间的相互转换,即实现了基本表与存储文件之间的相互转换,从而保证了数据的物理独立性。

试题答案

(39) C

试题 24(2012 年上半年试题 40-44)

某公司销售数据库的商品、仓库关系模式及函数依赖集 F1、F2 如下:

商品 (商品号, 商品名称, 生产商, 单价), $F1 = \{\text{商品号} \rightarrow \text{商品名称}, \text{商品号} \rightarrow \text{生产商}, \text{商品号} \rightarrow \text{单价}\}$, 商品关系的主键是 ()。仓库 (仓库号, 地址, 电话, 商品号, 库存量), $F2 = \{\text{仓库号} \rightarrow (\text{地址}, \text{电话}), (\text{仓库号}, \text{商品号}) \rightarrow \text{库存量}\}$ 。仓库关系的主键是 (), 外键是 ()。

仓库关系模式 (), 为了解决这一问题, 需要将仓库关系分解为 ()。

- A.商品号
- B.商品号, 商品名称
- C.商品号, 生产商
- D.商品名称, 生产商

- A.仓库号
- B.仓库号，商品号
- C.仓库号，电话
- D.地址，电话

- A.仓库号
- B.地址
- C.电话
- D.商品号

- A.存在冗余、插入异常和删除异常，以及修改操作的不一致
- B.不存在冗余，但存在插入异常和删除异常
- C.不存在修改操作的不一致，但存在冗余和插入异常
- D.不存在冗余、插入异常，但存在删除异常和修改操作的不一致

- A.仓库 1（仓库号，地址）和仓库 2（仓库号，电话，商品号，库存量）
- B.仓库 1（仓库号，地址，电话）和仓库 2（商品号，库存量）
- C.仓库 1（仓库号，电话）和仓库 2（仓库号，地址，商品号，库存量）
- D.仓库 1（仓库号，地址，电话）和仓库 2（仓库号，商品号，库存量）

试题分析

本题考查应试者对关系模式中主键、外键和模式分解及相关知识的掌握程度。

从商品关系的函数依赖集 F1 可以导出商品号决定商品关系的全属性，所以商品

号是商品关系的主键。

从仓库关系的函数依赖集 F2 可以导出 (仓库号, 商品号) 决定仓库关系的全属性, 所以仓库关系的主键是 (仓库号, 商品号)。又由于商品号是商品关系的主键, 故商品号是仓库关系的外键。

仓库关系存在冗余、插入异常和删除异常, 以及修改操作的不一致。例如, 仓库号为“12”的商品有 3 种, 其地址就要重复 3 次, 如下表所示, 故仓库关系存在冗余

仓库关系				
仓库号	地址	电话	商品号	库存量
12	高新路 6 号	8601	100024	168
12	高新路 6 号	8601	100025	568
12	高新路 6 号	8601	230023	398
13	友谊路 6 号	8602	150004	2600
13	友谊路 6 号	8602	160005	1218
...

试题 (44) 正确的答案 D。为了解决仓库关系模式存在的问题需要进行模式分解, 其中, 选项 A 分解存在的问题是仓库 2 不属于第三范式, 因为存在非主属性对码的部分函数依赖, 即仓库号 \rightarrow 电话。选项 B 分解存在的问题是分解有损连接, 即分解的新关系模式仓库 1 和仓库 2 无法恢复到原关系。选项 C 分解存在的问题与 A 类同, 分析略。

选项 D 分解是即保持函数依赖, 又无损连接, 分解的结果如下:

仓库 1			仓库 2		
仓库号	地址	电话	仓库号	商品号	库存量
12	高新路 1 号	8601	12	100024	168
13	友谊路 6 号	8602	12	100025	568
			12	230023	398
			13	150004	2600
			13	160005	1218
		

试题答案

(40) A (41) B (42) D (43) A (44) D

试题 25(2012 年上半年试题 45)

若对关系 $R(A, B, C, D)$ 、 $S(C, D, E)$ 进行 $\pi_{1,2,3,4,7}(\sigma_{3=5 \wedge 4=6}(R \times S))$ 运算，则该关系代数表达式与 () 是等价的。

A. $R \bowtie S$

B. $\sigma_{3=5 \wedge 4=6}(\pi_{1,2,3,4,7}(R \times S))$

C. $\pi_{A,B,C,D,E}(R \times S)$

D. $\pi_{1,2,3,4,7}(\sigma_{3=5}(R) \times \sigma_{4=6}(S))$

试题分析

本题考查关系代数运算方面的基础知识。

自然联接 是一种特殊的等值连接，它要求两个关系中进行比较的分量必须是相同的属性组，并且在结果集中将重复属性列去掉。本试题中 $\sigma_{3=5 \wedge 4=6}(R \times S)$ 的含义是 $R \times S$ 后，选取 R 和 S 关系中 $R.C=S.C \wedge R.D=S.D$ 的元组，再进行 $R.A$ 、 $R.B$ 、 $R.C$ 、 $R.D$ 和 $S.E$ 的投影关系运算。可见该关系运算表达式与 $R \bowtie S$ 是等价的。

试题答案

(45) A

试题 26(2011 年上半年试题 40)

数据库的视图与基本表之间，基本表与存储文件之间分别通过建立 () 之间的映像，保证数据的逻辑独立性和物理独立性。

A. 模式到内模式和外模式到内模式

B. 外模式到内模式和内模式到模式

C.外模式到模式和模式到内模式

D.内模式到模式和模式到外模式

试题分析

本题考查数据库系统管理方面的基础知识。

数据库的三级模式结构，其中，视图对应外模式、基本表对应模式、存储文件对应内模式。数据库系统在三级模式之间提供了两级映像：模式/内模式映像、外模式/模式映像。正因为这两级映像保证了数据库中的数据具有较高的逻辑独立性和物理独立性。

① 外模式/模式的映像：存在于外部级和概念级之间，实现了外模式到概念模式之间的相互转换。数据的逻辑独立性是指用户的应用程序与数据库的逻辑结构是相互独立的。数据的逻辑结构发生变化后，用户程序也可以不修改。但是，为了保证应用程序能够正确执行，我们需要修改外模式/概念模式之间的映像。

② 模式/内模式的映像：存在于概念级和内部级之间，实现了概念模式到内模式之间的相互转换。数据的物理独立性是指当数据库的内模式发生改变时，数据的逻辑结构不变。由于应用程序处理的只是数据的逻辑结构，这样物理独立性可以保证，当数据的物理结构改变了，应用程序不用改变。但是，为了保证应用程序能够正确执行，我们需要修改概念模式/内模式之间的映像。

试题答案

(40) C

试题 27(2011 年上半年试题 41-42)

给定关系模式 $R(U, F)$ ，其中，属性集 $U = \{\text{城市, 街道, 邮政编码}\}$ ，函数依赖集 $F = \{(\text{城市, 街道}) \rightarrow \text{邮政编码}, \text{邮政编码} \rightarrow \text{城市}\}$ 。关系 $R()$ ，且分别有 $()$ 。

- A. 只有 1 个候选关键字 “城市, 街道”
- B. 只有 1 个候选关键字 “街道, 邮政编码”
- C. 有 2 个候选关键字 “城市, 街道” 和 “街道, 邮政编码”
- D. 有 2 个候选关键字 “城市, 街道” 和 “城市, 邮政编码”

- A. 1 个非主属性和 2 个主属性
- B. 0 个非主属性和 3 个主属性
- C. 2 个非主属性和 1 个主属性
- D. 3 个非主属性和 0 个主属性

试题分析

本题考查关系数据库规范化理论方面的基础知识。

试题 (41) 的正确答案是 C。因为根据函数依赖定义，可推出 $(\text{城市, 街道}) \rightarrow U$ ， $(\text{邮政编码, 街道}) \rightarrow U$ ，所以 “城市, 街道” 和 “街道, 邮政编码” 为候选关键字。

试题 (42) 的正确答案是 B。因为根据主属性的定义，“包含在任何一个候选码中的属性叫做**主属性**(Prime attribute)，否则叫做**非主属性**(Nonprime attribute)”，所以关系 中的 3 个属性都是主属性，而无非主属性。

试题答案

(41) C (42) B

试题 28(2011 年上半年试题 43-45)

设有员工实体 Employee（员工号，姓名，性别，年龄，电话，家庭住址，家庭成员，关系，联系电话）。其中，“家庭住址”包括邮编、省、市、街道信息；

“家庭成员，关系，联系电话”分别记录了员工亲属的姓名、与员工的关系以及联系电话，且一个员工允许有多个家庭成员。

员工实体 Employee 的主键为（ ）；该关系属于（ ）；为使数据库模式设计更合理，对于员工关系模式 Employee（ ）。

- A. 员工号
- B. 员工号，家庭成员
- C. 姓名
- D. 姓名，家庭成员

- A. 2NF，无冗余，无插入异常和删除异常
- B. 2NF，无冗余，但存在插入异常和删除异常
- C. 1NF，存在冗余，但不存在修改操作的不一致
- D. 1NF，存在冗余和修改操作的不一致，以及插入异常和删除异常

- A. 只允许记录一个亲属的姓名、与员工的关系以及联系电话
- B. 可以不作任何处理，因为该关系模式达到了 3NF
- C. 增加多个家庭成员、关系及联系电话字段
- D. 应该将家庭成员、关系及联系电话加上员工号作为一个独立的模式

试题分析

本题考查关系数据库方面的基础知识。

试题 (43) 的正确答案为 B。因为一个员工可以有多个家庭成员，故为了唯一区分 Employee 关系中的每一个元组，其主键为 (员工号，家庭成员)。

试题 (44) 的正确答案为 D。关系模式 $Employee \in 1NF$ ，原因是员工号 \rightarrow (姓名，性别，年龄，电话，家庭住址)，即非主属性 (姓名，性别，年龄，电话，家庭住址) 不完全依赖于码 “员工号，家庭成员”，故 Employee 不属于 2NF。1NF 存在四个问题：冗余度大、引起修改操作的不一致性、插入异常和删除异常。

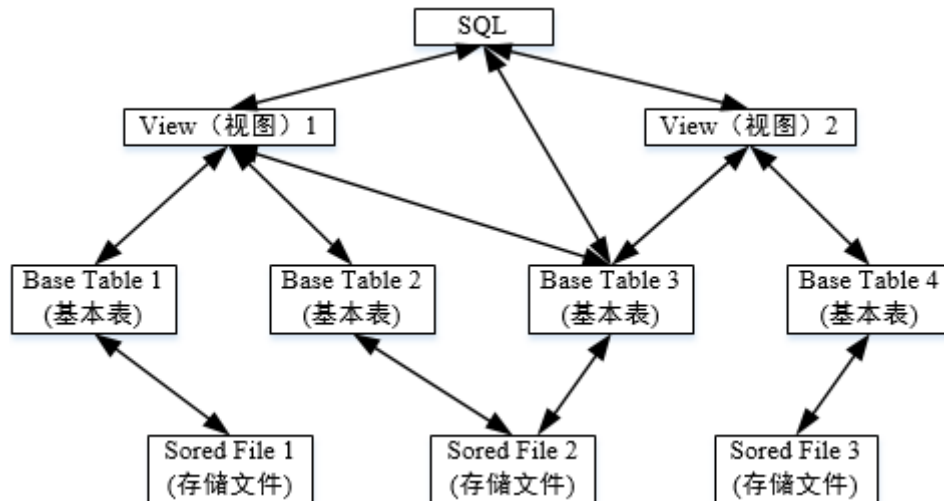
试题 (45) 的正确答案为 D。如果某员工有 5 个亲属，那么该员工关系中 “员工号，姓名，性别，年龄，电话，家庭住址” 将重复出现 5 次，为了将数据库模式设计的更合理，应该消除冗余，即将家庭成员、关系及联系电话加上员工号设计成为一个独立的模式。

试题答案

(43) B (44) D (45) D

试题 29(2010 年上半年试题 40)

SQL 语言支持关系数据库的三级模式结构图如下所示，图中视图、基本表、存储文件分别对应 ()。



- A.模式、内模式、外模式
- B.外模式、模式、内模式
- C.模式、外模式、内模式
- D.外模式、内模式、模式

试题分析

SQL 语言支持关系数据库的三级模式结构，其中：视图对应外模式、基本表对应模式、存储文件对应内模式。

试题答案

(40) B

试题 30(2010 年上半年试题 41-42)

确定系统边界应在数据库设计的（ ）阶段进行；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

A.需求分析

B.概念设计

C.逻辑设计

D.物理设计

试题分析

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、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，确定系统边界、收集支持系统目标的基础数据及其处理方法。

逻辑设计阶段的任务之一是对关系模式进一步地规范化处理。因为生成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，这就需要根据规范化理论对关系模式分解之后，消除冗余和更新异常。不过有时根据处理要求，可能还需要增加部分冗余以满足处理要求。逻辑设计阶段的任务就需要作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。

试题答案

(41) A (42) C

试题 31(2010 年上半年试题 43-45)

某销售公司数据库的零件关系（零件号，零件名称，供应商，供应商所在地，库存量）如下表所示，其中同一种零件可由不同的供应商供应，一个供应商可以供应多种零件。零件关系的主键为（ ），该关系存在冗余以及插入异常和删除异

常等问题。为了解决这一问题需要将零件关系分解为（ ），分解后的关系模式可以达到（ ）。

零件号	零件名称	供应商	供应商所在地	库存量
010023	P2	S1	北京市海淀区 58 号	380
010024	P3	S1	北京市海淀区 58 号	1350
010022	P1	S2	西安市雁塔区 2 号	160
010023	P2	S2	西安市雁塔区 2 号	1280
010024	P3	S2	西安市雁塔区 2 号	3900
010022	P1	S3	北京市新城区 65 号	2860
...

(43) A.零件号，零件名称

B.零件号，供应商

C.零件号，供应商所在地

D.供应商，供应商所在地

(44) A.零件 1（零件号，零件名称，供应商，供应商所在地，库存量）

B.零件 1（零件号，零件名称）、零件 2（供应商，供应商所在地，库存量）

C.零件 1（零件号，零件名称）、零件 2（零件号，供应商，库存量）、零件 3（供应商，供应商所在地）

D.零件 1（零件号，零件名称）、零件 2（零件号，库存量）、零件 3（供应商，供应商所在地）、零件 4（供应商所在地，库存量）

(45) A.1NF

B.2NF

C.3NF

D.4NF

试题分析

试题（43）的正确选项为 B。根据题意，零件关系的主键为（零件号，供应商）。

试题（44）的正确选项为 C。因为关系 P 存在冗余以及插入异常和删除异常等问题。

为了解决这一问题需要将零件关系分解。选项 A，选项 B 和选项 D 是有损连接的，且不保持函数依赖性故分解是错误的，例如，分解为选项 A、选项 B 和选项 D 后，用户无法查询某零件由哪些供应商供应，原因是分解有损连接的，且不保持函数依赖。

试题（45）的正确选项为 C。因为，原零件关系存在非主属性对码的部分函数依赖：（零件号，供应商）→ 供应商所在地，但是供应商 → 供应商所在地，故原关系模式零件非 2NF 的。分解后的关系模式零件 1、零件 2 和零件 3 消除了非主属性对码的部分函数依赖，同时不存在传递依赖，故达到 3NF。

试题答案

(43) B (44) C (45) C