

第 3 章

数据库基础知识

题型 1：基本概念

命题方向：该考点主要考查数据库和数据库管理系统的概念、数据库管理技术的发展。

【典型题 1】（2005 年上半年上午试题 20）

数据库管理技术是在____(1)____的基础上发展起来的。

- (1) A. 文件系统 B. 编译系统 C. 应用程序系统 D. 数据库管理系统

分析：

数据管理是数据处理的中心问题，是对数据进行分类、组织、编码、存储、检索和维护。数据管理技术的发展经历了人工管理、文件系统和数据库系统阶段。数据库系统是在文件系统的基础上发展起来的，实现了大量关联数据有组织地、动态地存储，方便多用户访问，与文件系统的重要区别是数据的充分共享、交叉访问，以及与应用程序的高度独立。

答案：(1) A

【典型题 2】（2002 年上午试题 31）

相对于数据库系统，文件系统的主要缺陷有数据联系弱、数据的不一致性和数据的____(1)____。

- (1) A. 可重用性差 B. 安全性差 C. 非持久性 D. 冗余性

分析：

文件系统解决了应用程序和数据之间的一个公共接口问题，使应用程序可采用统一存取方法来操作数据。与数据库系统相比，文件系统具有数据冗余度大、数据不一致和数据联系弱等缺点。

数据冗余度大是指文件与应用程序密切相关，相同的数据经常被重复定义、重复存储。不同的应用程序不能共享相同的数据，必然导致冗余。由于相同数据的重复存储、单独管理，给数据的修改和维护带来难度，容易造成数据不一致。文件系统中数据组织成记录，记录由字段组成，记录内部有了一定的结构，但是文件之间是孤立的，从整体上看没有反映现实世界事物之间的内在联系，数据联系弱。

答案：(1) D

考点链接 3.1.1: 数据库管理技术的发展

数据处理是对各种数据进行收集、存储、加工和传播的一系列活动。数据管理是数据处理的中心问题，是对数据进行分类、组织、编码、存储、检索和维护。数据管理技术的发展经历了人工管理、文件系统和数据库系统阶段。

(1) 人工管理阶段

早期的数据处理都是通过手工进行的，当时的计算机主要用于科学计算，没有专门管理数据的软件。数据处理具有以下的特点：数据量较少，数据和程序一一对应，一组数据对应一个程序，数据面向应用，独立性很差；数据不保存；没有软件对数据进行管理，数据对程序不具有独立性，一旦数据在存储器上改变物理地址，就需要改变相应的用户程序；程序对数据的依赖性太强；数据组之间可能有许多重复的数据，造成数据冗余。

(2) 文件系统阶段

在文件系统中，按一定的规则将数据组织成一个文件，应用程序通过文件系统对文件中的数据进行存取和加工。

文件系统的最大特点是解决了应用程序和数据之间的一个公共接口问题，使应用程序可采用统一存取的方法来操作数据。在文件系统阶段数据管理的特点：数据可以长期保存，数据的逻辑结构和物理结构有了区别，程序可以按名访问，不必关心数据的物理位置；数据不属于某个特定的应用，可以重复使用；文件组织形式多样化，有索引文件、链接文件和 Hash 文件等，当文件之间相互独立时，数据间的联系要通过程序去构造。

文件系统还存在着很大的缺点。

①数据冗余度大：文件与应用程序密切相关，相同的数据经常被重新重复定义、重复存储。不同的应用程序不能共享相同的数据，必然导致冗余。

②数据不一致性：由于相同数据的重复存储、单独管理，给数据的修改和维护带来难度，容易造成数据不一致。

③数据联系弱：文件系统中数据组织成记录，记录由字段组成，记录内部有了一定的结构，但是文件之间是孤立的，从整体上看没有反映现实世界事物之间的内在联系，因此很难对数据进行合理的组织以适应不同应用的需要。

(3) 数据库系统阶段

数据库系统是由计算机软件、硬件资源组成的系统，它实现了大量数据有组织地、动态地存储，方便多用户访问。

数据库系统阶段数据管理的特点：采用复杂的数据模型表示数据结构，数据不再面向某个应用，而是面向整个应用系统，数据冗余明显减少，实现了数据共享；有较高的独立性。

数据库系统与文件系统的区别：数据库对数据的存储是按照同一结构进行的，不同的应用程序可以直接操作这些数据。数据库对数据的完整性、惟一性和安全性都提供了一套有效的管理手段。

【典型题 3】 (2005 年上半年上午试题 22)

建立数据库系统的主要目标是减少数据冗余，提高数据的独立性，并集中检查 (1)。

- (1) A. 数据操作性 B. 数据兼容性 C. 数据完整性 D. 数据可维护性

分析:

数据库系统管理数据的特点如下。

- (1) 数据共享性高、冗余度低且易扩充,这是数据库系统最大的改进。
- (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) 数据库管理系统提供了4个方面的数据控制功能,即数据完整性、数据安全性、数据库的并发控制和数据库的恢复。

针对本道题目,建立数据库系统的主要目标是减少数据冗余,提高数据的独立性,并集中检查数据完整性。

答案:(1) C

【典型题4】 (2004年上半年上午试题17)

数据库系统从广义的角度看,应该由(1)组成。

- | | |
|---------------------|----------------------|
| (1) A. 数据库、硬件、软件和人员 | B. 数据库、硬件、数据库管理系统和软件 |
| C. 数据库、软件和人员 | D. 数据库、数据库管理系统和人员 |

分析:

数据库系统广义上看是由数据库、硬件、软件和人员组成的。数据库是指长期存在于计算机内的、有组织的并且可共享的数据集合。软件包括操作系统、数据库管理系统及应用程序,其中数据库管理系统是数据库系统的核心软件。人员有4类:系统分析员和数据库设计人员、应用程序员、数据库管理员与最终用户。

答案:(1) A

【典型题5】 (2001年上午试题35)

数据库技术中的“脏数据”是指(1)的数据。

- | | |
|-----------|---------------|
| (1) A. 错误 | B. 回返 |
| C. 未提交 | D. 未提交的随后又被撤销 |

分析:

数据库存储了错误的数据或者用户读了不正确的数据,即未提交的随后又被撤销的数据为“脏数据”。

答案:(1) D

考点链接 3.1.2: 数据库系统的组成与特点

数据是描述事物的符号记录,具有多种表现形式,可以是文字、图形、图像和声音等。

信息是现实世界事物的存在方式或状态的反映,具有可感知、可存储、可加工、可传递和可再生等自然属性。

数据库系统广义上看是由数据库、硬件、软件和人员组成。

数据库是指长期存储在计算机内的,有组织的,可共享的数据集合。数据库中的数据按一

定的数学模型组织、描述和储存，具有较小的冗余度，较高的数据独立性和易扩展性，并可为各种用户共享。

硬件是构成计算机系统的各种物理设备，包括存储数据所需的外部设备。

软件包括操作系统、数据库管理系统及应用程序。数据库管理系统是数据库的核心软件，要在操作系统的支持下工作，解决如何科学地组织和存储数据，如何高效地获取和维护数据的系统软件。

人员主要有4类。第一类为系统分析员和数据库设计人员，系统分析员负责应用系统的需求分析和规范说明；数据库设计人员负责数据库中数据的确定、数据库各级模式的设计。第二类为应用程序员，负责编写使用数据库的应用程序，这些应用程序可对数据进行检索、建立、删除和修改。第三类为最终用户。第四类是数据库管理员，负责数据库的总体信息控制，具体职责包括决定数据库中的信息内容和结构、决定数据库的存储结构和存取策略、定义数据库的安全性和完整性约束条件、监控数据库的使用和运行、数据库的性能改进、数据库的重组和重构。

题型 2：数据模型

命题方向：该考点主要考查数据模型的基本概念、数据模型的三要素、E-R模型和几种基本的数据模型。

【典型题 6】（2006 年上半年上午试题 54）

数据模型的三要素包括___(1)___。

- (1) A. 外模式、模式、内模式 B. 网状模型、层次模型、关系模型
C. 实体、联系、属性 D. 数据结构、数据操纵、完整性约束

分析：

数据模型的三要素是数据结构、数据操作和数据的约束条件。数据结构是所研究的对象类型的集合，是对系统静态特性的描述。数据操作是数据库中对象的实例允许执行的操作的集合，包括操作和操作规则。数据的约束条件是一组完整性规则的集合。

答案：(1) D

考点链接 3.2.1：数据模型基本概念与三要素

(1) 数据模型的基本概念

信息世界是现实世界在人们头脑中的反映，数据库常用的术语是实体、实体集、属性和码。机器世界是按计算机的观点对数据建模，描述数据的术语有记录、文件、字段和记录码。信息世界和机器世界相关术语的对应关系如下。

属性记录实体某方面的特性，字段标记实体属性的命名单位。实体表示客观存在并能相互区别的事物；记录是字段的有序集合，一般情况下，一条记录描述一个实体。码是惟一区分实体的属性或属性集，记录码是惟一标识文件中的每条记录的字段或字段集。实体集是具有共同特性的实体的集合，文件是统一类记录的汇编。实体型是属性的集合，记录型是记录的结构定义。

(2) 数据模型的三要素

数据模型的三要素是数据结构、数据操作和数据的约束条件。

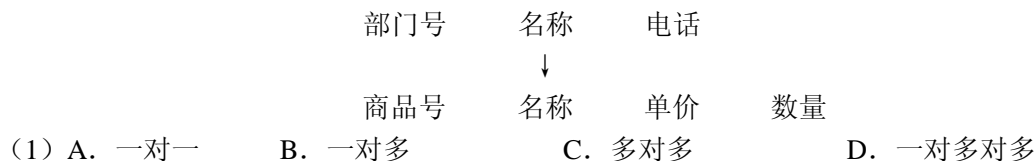
数据结构是对系统静态特性的描述，是所研究的对象类型的集合。

数据操作是对系统动态特性的描述，包括操作和操作规则，如检索、插入和修改等。

数据的约束条件是一组完整性规则的集合。完整性规则提供了一种手段来保证当授权用户对数据库作修改时不会破坏数据的一致性，可防止外来因素对数据的恶意破坏。

【典型题 7】 (2005 年下半年上午试题 54)

某商场的部门和商品两个实体之间的关系如下图所示。假设每个部门负责销售若干种商品，每种商品只能由一个部门负责销售，那么部门和商品之间存在着(1)的联系。



分析：

两个不同实体集之间存在一对一、一对多和多对多的联系类型。

一对一是指实体集 E1 中一个实体最多只与实体集 E2 中的一个实体相联系。

一对多是指实体集 E1 中一个实体可与实体集 E2 中的多个实体相联系。

多对多是指实体集 E1 中多个实体可与实体集 E2 中的多个实体相联系。

在部门和商场两个实体集之间，一个部门可以与多种商品相联系，而一种商品只与一个部门相联系，显然这种联系为一对多的联系类型。

答案：(1) B

考点链接 3.2.2：常用的数据模型

常用的数据模型分为概念数据模型和基本数据模型。

(1) 概念数据模型

概念数据模型也称信息模型，是按用户的观点对数据和信息建模，是现实世界到信息世界的第二层抽象，主要用于数据库设计。这类模型中最著名的是实体联系模型，即 E-R 模型。实体、属性和联系是 E-R 模型的基本概念。

实体指表示现实世界中存在并能区别的事物。在 E-R 图中，实体用矩形框表示。

属性是实体某方面的特性。在 E-R 图中，属性用椭圆表示。

联系一般指描述实体集之间的依赖关系。两个不同实体集之间存在一对一、一对多和多对多的联系类型。在 E-R 图中，联系用菱形表示。

一对一：指实体集 E1 中一个实体最多只与实体集 E2 中的一个实体相联系，记为 1:1。

一对多：指实体集 E1 中一个实体可与实体集 E2 中的多个实体相联系，记为 1:n。

多对多：指实体集 E1 中多个实体可与实体集 E2 中的多个实体相联系，记为 m:n。

(2) 基本的数据模型

基本数据模型是按计算机系统的观点对数据建模，是现实世界数据特征的抽象，用于

DBMS 的实现。基本的数据模型有层次模型、网状模型和关系模型。

①层次模型采用树型结构表示数据与数据之间的关系。每一个节点表示一个记录类型，根节点以外的其他节点有且仅有一个双亲节点，上层和下一层类型之间的联系是 1:m。

②网状结构是采用网络结构表示实体类型及实体间联系的数据模型，允许一个以上的节点无双亲，一个节点可以有多个双亲。

③关系模型用二维表格结构来表示实体及其联系，一个二维表就叫做一个关系。二维表中的一列称为属性，一行称为元组，其值能唯一地标志元组的一个或多个属性称为主码或关键字，元组的一个属性值称为分量，属性的取值范围称为域。

关系模型是规范化的，概念单一，存储路径对用户是透明的，具有更好的数据独立性和安全保密性，与网状模型、层次模型的最大差别是用主码而不是用指针导航数据。

题型 3: DBMS 的功能和特征

命题方向：该考点主要考查数据库管理系统的功能、特征。

【典型题 8】（2004 年上半年上午试题 18）

数据字典中存放__（1）__，数据库管理系统对应用程序的操作都要通过数据字典来进行。

- (1) A. 数据库管理系统软件 B. 数据定义语言 DDL
C. 数据操纵语言 DML D. 数据库体系结构的描述

分析：

数据字典通常包括数据项、数据结构、数据流、数据存储和处理 5 个部分。其中，数据项是数据的最小组成单位，若干个数据项可以组成一个数据结构，数据字典通过对数据项和数据结构的定义来描述数据流、数据存储的逻辑内容。数据字典是数据库各种数据描述的集合，即数据库体系结构的描述。

答案：(1) D

【典型题 9】（2003 年上午试题 21）

支持数据库各种操作的软件系统叫做__（1）__。

- (1) A. 数据库管理系统 B. 文件系统
C. 数据库系统 D. 操作系统

分析：

数据库管理系统向用户提供数据操纵语言，实现对数据库中数据的基本操作，如检索、插入、修改和删除等。

答案：(1) A

【典型题 10】（2000 年上午试题 7）

DBMS 是位于__（1）__之间的一层数据管理软件。

答案：（1）C

考点链接 3.3.2：数据库管理系统的特征

通过数据库管理系统管理数据的特点如下。

（1）数据结构化且统一管理

数据库系统采用了复杂的数据模型表示数据结构，这不仅可以描述数据本身的特点，还描述数据之间的关系。

数据不再面向某个应用，而是面向整个应用系统。数据易维护、易扩展，数据冗余明显减少，真正实现了数据共享。

（2）有较高的数据独立性

数据的存储由数据库管理系统来负责，应用程序关心的只是数据的逻辑结构，无需了解数据在磁盘上的数据库中的存储形式，从而简化应用程序，大大减少应用程序编制的工作量。

（3）数据控制功能

数据库管理系统提供了数据控制功能，以适应共享数据的环境。数据控制功能包括以下几个方面。

数据库的安全性保护：保护数据库以防止不合法的使用所造成的数据泄露、更改或破坏，用户只能按规定对数据进行处理。

数据的完整性：指数据库的正确性和相容性，防止合法用户使用数据库向数据库加入不符合语义的数据，保证数据库中数据的正确性，避免非法的更新。

并发控制：在多用户共享的系统中，许多用户可能同时对同一数据进行操作。并发操作带来的问题是数据的不一致性，主要有 3 类：丢失更新、不可重复读和读脏数据。造成这些问题的主要原因是事务的兵法操作破坏了事务的隔离性。数据库管理系统的并发控制子系统负责协调并发事务的执行，保证数据库的完整性不受破坏。

故障恢复：数据库中的 4 类故障分别是事务内部故障、系统故障、介质故障及计算机病毒。故障恢复主要是指恢复数据库本身，即在故障引起数据库当前状态不一致后，将数据库恢复到某个正确状态或一致状态。

题型 4：数据库系统体系结构

命题方向：该考点主要考查数据库的三级模式结构、两级映射及用户角度的数据库体系结构的分类与特点。

【典型题 12】（2000 年上午试题 6）

数据库的三级体系结构即关系子模式、关系模式与存储模式，是对 （1） 抽象的 3 个级别。

- （1）A. 存储器 B. 数据库系统 C. 数据 D. 数据库管理系统

分析：

数据库系统设计人员在视图层、逻辑层和物理层对数据抽象，通过外模式、概念模式和内

据的存储，从而简化应用程序。数据的独立性是由 DBMS 的两级映射功能来保证的。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。当数据的物理结构发生了变化，只需改变模式/内模式映像，而使概念模式尽可能保持不变。也就是说数据的内模式改变，而模式可以不变，因而应用程序也不用改变，达到了数据的物理独立性。

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，用户程序可以不改变。如果数据库的概念模式要进行修改，那么外模式/模式映像也要做相应的修改，但外模式很可能保持不变，即达到了数据的逻辑独立性。

题型 5：关系数据库

命题方向：该考点主要考查关系数据库的基本概念、关系数据库模式和完整性约束。

【典型题 14】（2004 年下半年上午试题 20）

关系数据库是___(1)___的集合，它由一个或多个关系模式定义。

- (1) A. 元组 B. 列 C. 字段 D. 表

分析：

关系数据库系统采用关系模式作为数据的组织方式。在关系模型中用表格表达实体集，以及实体集之间的联系。关系是一个实例，也是一张表，对应于程序设计语言中的变量的概念。

答案：(1) B

【典型题 15】（2003 年上午试题 20）

对一个数据库视图进行查询时，就是从___(1)___中导出数据。

- (1) A. 一个或若干个基本表 B. 一个或若干个索引文件
C. 一个或若干个视图 D. 一个视图

分析：

视图是由基本表或其他视图表导出的表，它本身不独立存储在数据库中，数据库中只存放它的定义，当对一个数据库视图进行查询时，就是从一个或几个基本表中导出数据。

答案：(1) A

考点链接 3.5.1：关系数据库的基本概念

关系数据库系统采用关系模型作为数据的组织形式，用表格结构表达实体集，以及实体集之间的联系，其最大特色是描述的一致性。关系模型是由若干个关系模式组成的集合，一个关系模式相当于一个记录型，对应于程序设计语言中类型定义的概念。

(1) 属性和域

属性是描述一个事物的若干特征。每个属性的取值范围对应一个值的集合，成为该属性的域。一般在关系数据模型中，限制所有的域都应是原子数据，要求每一个属性都不可再分。关

系数据模型的这种限制称为第一范式条件。

(2) 关系的相关名词

候选码: 若关系中的某一属性或属性组的值能惟一地标志一个元组, 则称该属性或属性组为候选码。

主码: 若一个关系有多个候选码, 则选定其中一个为主码。

主属性: 包含在任何候选码中的属性称为主属性。不包含在任何候选码中的属性称为非主属性。

外码: 如果关系模式 R 中的属性或属性组不是该关系的码, 但它是其他关系的码, 那么该属性集对关系模式 R 而言是外码。

全码: 关系模型的所有属性组是这个关系的候选码, 称为全码。

(3) 关系的性质

一个基本关系具有以下 5 条性质。

- ①分量必须取原子值, 每一个分量必须是不可再分的数据项。
- ②列是同质的, 每一列中的分量是同一类型的数据, 来自同一个域。
- ③属性不能重名, 每列为一个属性, 不同的列可来自同一个域。
- ④行列的顺序无关。
- ⑤任何两个元组不能完全相同, 这是由主码约束来保证的。

(4) 关系的 3 种类型

基本表: 实际存在的表, 是实际存储数据的逻辑表示。

查询表: 查询结果对应的表。

视图表: 是由基本表或其他视图表导出的表。由于它本身不独立存储在数据库中, 数据库中只存放它的定义, 所以常称为虚表。

【典型题 16】 (2002 年上午试题 39)

给定关系模式学生(学号, 课程号, 名次), 若每一名学生每门课程有一定的名次, 每门课程每一名次只有一名学号, 则以下叙述中错误的是 (1)。

- (1) A. (学号, 课程号) 和 (课程号, 名次) 都可以作为候选键
- B. 只有 (学号, 课程号) 能作为候选键
- C. 关系模式属于第三范式
- D. 关系模式属于 BCNF

分析:

根据题目的陈述可以得到以下的函数依赖。

(学号, 课程号) \rightarrow 名次, (课程号, 名次) \rightarrow 学号

所以 (学号, 课程号) 和 (课程号, 名次) 都可以作为候选键。另外, 由于在学生关系中无非主属性, 且每一个决定因素都包含码, 故该学生关系属于 BCNF。BCNF 的规范化程度比第三范式要高, 因此该关系模式也属于第三范式。

答案: (1) B

【典型题 17】 （2001 年上午试题 34）

在关系模式 $R(U)$ 中，如果 $X \rightarrow Y$ 和 $X \rightarrow Z$ 成立，则 $X \rightarrow YZ$ 也成立，这条规则称为 (1) 。

- (1) A. 自反律 B. 增广律 C. 合并律 D. 分解律

分析：

函数依赖指属性或属性组之间的对应关系，是一种完整性约束形式，语义约束，因而是模式的一部分。若对 $R(U)$ 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 上属性值不等，则称“ X 函数决定 Y ”，或“称 Y 函数依赖于 X ”，记作 $X \rightarrow Y$ ， X 称为决定元素。

函数依赖满足如下关系。

自反率：如果 B 是 A 的子集，则 $A \rightarrow B$ （平凡依赖）。

增广率：如果 $A \rightarrow B$ ，则 $AC \rightarrow BC$ 。

传递率：若 $A \rightarrow B$ 且 $B \rightarrow C$ ，则 $A \rightarrow C$ 。

上述 3 个规则是 3 条基本定理，可以得到如下推论。

自含： $A \rightarrow A$ 。

分解： $A \rightarrow BC$ ，则 $A \rightarrow B$ ，且 $A \rightarrow C$ 。

合并： $A \rightarrow B$ ，且 $A \rightarrow C$ ，则 $A \rightarrow BC$ 。

复合： $A \rightarrow B$ ，且 $C \rightarrow D$ ，则 $AC \rightarrow BD$ 。

答案：(1) C

【典型题 18】 （2000 年上午试题 7）

关系数据库的数据及更新操作必须遵循 (1) 等完整性规则。

- (1) A. 实体完整性和参照完整性
 B. 参照完整性和用户定义的完整性
 C. 实体完整性和用户定义的完整性
 D. 实体完整性、参照完整性和用户定义的完整性

分析：

完整性规则提供了一种手段来保证当授权用户对数据库作修改时不会破坏数据的一致性。关系模型的完整性规则是对关系的某种约束条件。关系的完整性共分为 3 类：实体完整性、参照完整性和用户定义完整性。关系数据库的数据及更新操作必须遵循这 3 类完整性规则。

答案：(1) D

考点链接 3.5.2：关系数据库的模式及完整性约束

(1) 关系数据库模式

关系数据库模式是关系数据库结构的描述，它包括若干域的定义以及在这些域上定义的若干关系模式。关系模式对应于程序设计语言中类型定义的概念。

关系的描述称为关系模式，可以形式化地表示为 $R(U,D,dom,F)$ 。其中 R 表示关系名， U 是组成该关系的属性名集合， D 是属性的域， dom 是属性向域的映像集合， F 为属性间数据的函

数依赖关系集合。通常关系模式简记为 $R(U)$ 。

(2) 函数依赖

若对 $R(U)$ 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 上属性值不等, 则称“ X 函数决定 Y ”, 或“称 Y 函数依赖于 X ”, 记作 $X \rightarrow Y$, X 称为决定元素。

在 $R(U)$ 中, 如果 $X \rightarrow Y$, 并且对于 X 的任何一个真子集 X' 都不能函数决定 Y , 则称 Y 完全函数依赖于 X , 记作 $X \xrightarrow{f} Y$ 。

在 $R(U)$ 中, 如果 $X \rightarrow Y$, 但不完全依赖于 X , 则称 Y 对 X 部分函数依赖, 记作 $X \xrightarrow{p} Y$ 。

函数依赖满足以下关系。

自反率: 如果 B 是 A 的子集, 则 $A \rightarrow B$ (平凡依赖)。

增广率: 如果 $A \rightarrow B$, 则 $AC \rightarrow BC$ 。

传递率: 若 $A \rightarrow B$ 且 $B \rightarrow C$, 则 $A \rightarrow C$ 。

上述 3 个规则是 3 条基本定理, 可以得到如下推论。

自含: $A \rightarrow A$ 。

分解: $A \rightarrow BC$, 则 $A \rightarrow B$, 且 $A \rightarrow C$ 。

合并: $A \rightarrow B$, 且 $A \rightarrow C$, 则 $A \rightarrow BC$ 。

复合: $A \rightarrow B$, 且 $C \rightarrow D$, 则 $AC \rightarrow BD$ 。

(3) 关系模式的规范化

第一范式: 如果关系模式 R 中每一属性都是不可分的基本数据项, 则 R 1NF。第一范式是对关系模式的最起码的要求, 不满足第一范式的数据库模式不能称为关系数据库。

第二范式: 如果关系模式 R 1NF, 并且每一个非主属性都完全函数依赖于关键码, 则 R 2NF。

第三范式: 如果关系模式 R 2NF, 且不存在非主属性传递依赖于其他的非关键字属性, 则 R 3NF。

BCNF 范式: 又称改进的第三范式, 如果关系模式 R 3NF, 并且没有一个非主属性是传递函数依赖于其他的非关键字属性, 则 R BCNF。

第四范式: 如果关系模式 R 3NF, 且没有非平凡且非函数依赖的多值依赖, 则 R 4NF。

以上几种范式之间, 4NF 隐含 BCNF, 而 BCNF 又隐含 3NF。即如果一组元组满足 4NF 条件, 那么它必然满足其他两个范式条件; 而且如果它满足 BCNF 条件, 那么它必然属于 3NF。

在关系数据库中, 对关系模式的基本要求是满足第一范式, 寻求解决关系模式存在的插入、删除异常, 数据冗余, 以及修改等复杂的问题, 这就是规范化的目的。

(4) 完整性约束

完整性规则是保证数据一致性的一种重要手段, 可防止对数据的恶意破坏。关系模型的完整性规则是对关系的某种约束条件, 关系的完整性约束分为 3 类: 实体完整性、参照完整性和用户定义完整性。

① 实体完整性: 指基本关系的主属性的不能去空值。

② 参照完整性: 又称实体完整性, 指不允许引用不存在的实体。参照完整性规定: 若 F 是基本关系 R 的外码, 它与基本关系 S 的主码 K_S 相对应, 则对于 R 中每个元组在 F 上的值必须

或者取空值，或者等于 S 中某个元组的主码值。

③用户定义的完整性：是针对某一具体的关系数据库的约束条件，反映某一具体应用所涉及的数据必须满足的语义要求，由应用的环境决定。

题型 6：关系代数运算

命题方向：该考点主要考查并、差、笛卡儿积、投影和选择 5 种基本关系代数运算，同时还有交、连接及除法运算。

【典型题 19】（2006 年上半年上午试题 57）

设有关系 R 、 S 和 T ，如下图所示，其中，关系 T 是 (1) 运算的结果。

A	B	C
a	b	c
d	a	f
c	b	d
a	b	p
d	a	m

A	B	C
a	b	i
d	a	f
d	a	p
c	b	d

A	B	D
a	b	i
d	a	p

- (1) A. $S \cup R$ B. $S - R$ C. $S \cap R$ D. $S \times R$

分析：

关系 R 与 S 的并是由属于 R 或属于 S 的元组构成的集合。关系 R 与 S 的差是由属于 R 但不属于 S 的元组构成的集合。关系 R 与 S 的交是由属于 R 同时又属于 S 的元组构成的集合。两个元数分别为 n 目和 m 目的关系 R 和 S 的广义笛卡儿积是一个 $(n+m)$ 列的元组的集合。

题目中关系 T 中的元组是由属于 S 但不属于 R 元组构成的集合，因此 $T=S-R$ 。

答案：(1) B

【典型题 20】（2005 年下半年上午试题 15）

在关系代数运算中，若对表 1 按 (1)，可以得到表 2。

商品号	名称	单价：元
010202010	牙刷	3
010202011	毛巾	10
010202012	毛巾	8
010202013	铅笔	0.5
010202014	钢笔	8

名称	单价：元
毛巾	10

- (1) A. 商品号都进行选择运算
 B. 名称、单价进行投影运算
 C. 名称等于“毛巾”进行选择运算，再按名称、单价进行投影运算
 D. 商品号等于“01020211”进行选择运算，再按名称、单价进行投影运算

分析:

选择是对一个二维表按条件 F 进行横向选择，从关系 R 中选择满足给定条件的元组。投影运算是对一个二维表进行纵向筛选，从 R 中选择出若干属性列，组成新的关系。

题目中，通过选择运算选取“01020211，毛巾，10”这一个元组，按名称、单价进行投影运算就可以得到“毛巾，10”。

答案: (1) D

【典型题 21】 (2005 年下半年上午试题 45~46)

若课程域 $D1=\{\text{数据库, 操作系统, 面向对象}\}$, 教师域 $D2=\{\text{王强, 刘亦}\}$, 学生域 $D3=\{\text{刘海华, 张刚, 吴琨, 梁谦}\}$, 则 $D1 \times D2 \times D3$ (1)。在关系代数运算中, $D1 \times D2$ 的结果有 (2) 个元组。

- (1) A. 可以用一个二维表来表示 B. 不能用一个二维表来表示
 C. 不是一个集合也不是一个关系 D. 不是一个集合但它是一个关系
 (2) A. 3 B. 4 C. 5 D. 6

分析:

笛卡儿积的形式定义如下。

$$R \times S = \{(a_1, a_2, \dots, a_m, b_1, b_2, \dots, b_n) \mid (a_1, a_2, \dots, a_m) \in R, (b_1, b_2, \dots, b_n) \in S\}$$

两个关系的笛卡儿积的结果仍然是一个二维表。 $D1 \times D2 \times D3 = (D1 \times D2) \times D3 = D4 \times D3$, $D1 \times D2$ 的结果为一二维表, $D4$ 再和 $D3$ 进行笛卡儿积得到的仍然是一个二维表。

由笛卡儿积的形式定义可以看出, $D1 \times D2 = \{(\text{数据库, 王强}), (\text{数据库, 刘亦}), (\text{操作系统, 王强}), (\text{操作系统, 刘亦}), (\text{面向对象, 王强}), (\text{面向对象, 刘亦})\}$, 结果中有 6 个元组。

答案: (1) A (2) D

【典型题 22】 (2005 年上半年上午试题 23)

在关系代数运算中, (1) 运算结果的关系模式与原关系模式相同。

- (1) A. 并 B. 笛卡儿积 C. 投影 D. 自然连接

分析:

在关系代数中, 交、并、差是二元运算, 要求参与运算的两个关系的结构必须相同, 运算结果的结构与原关系模式的结构相同。笛卡儿积的结果是一个列为两个关系列数的和; 投影运算取的是关系的某些列; 自然连接要求两个关系中进行比较的分量必须是相同的属性组, 并去掉结果集中的重复属性。这 3 种运算结果的结构与原关系模式可以不同。

答案: (1) A

【典型题 23】 （2002 年上午试题 38）

若 4 元关系 R 为 $R(A, B, C, D)$ ，则___(1)___。

- (1) A. $\pi_{A,C}(R)$ 为取属性值为 $A、C$ 的两列组成新关系
- B. $\pi_{1,3}(R)$ 为取属性值为 $A、C$ 的两列组成新关系
- C. $\pi_{1,3}(R)$ 与 $\pi_{A,C}(R)$ 是等价的
- D. $\pi_{1,3}(R)$ 与 $\pi_{A,C}(R)$ 是不等价的

分析：

$\pi_{A,C}(R)$ 是投影运算，是对属性 $A、C$ 的投影。属性 $A、C$ 分别位于第一列和第三列， $\pi_{A,C}(R)$ 相当于在关系 R 上第一列和第三列上投影，故 $\pi_{1,3}(R)$ 与 $\pi_{A,C}(R)$ 是等价的。

答案：(1) C

【典型题 24】 （2000 年上午试题 6）

关系代数中的 q 联接操作由___(1)___操作组合而成。元组比较操作 $(c1,c2) \leq (d1,d2)$ ，其意义等价于___(2)___。

- (1) A. σ 和 π B. σ 和 \times C. $\pi、\sigma$ 和 \times D. π 和 \times
- (2) A. $(c1 \leq d1) \text{ OR } (c2 \leq d2)$ B. $(c1 \leq d1) \text{ OR } ((c1 = d1) \text{ AND } (c2 \leq d2))$
- C. $(c1 \leq d1) \text{ AND } (c2 \leq d2)$ D. $(c1 < d1) \text{ OR } ((c1 = d1) \text{ AND } (c2 \leq d2))$

分析：

q 连接是指从关系 R 和 S 的笛卡儿积中选取属性间满足一定条件的元组，也就是说 q 联接操作由笛卡儿积和选择两种操作组合而成。 σ 和 \times 分别表示选择、笛卡儿积运算符。

$(c1,c2) \leq (d1,d2)$ 操作为两个元组比较大小，若 $c1 < d1$ ，则结果 $(c1,c2) < (d1,d2)$ ；若 $c1 = d1$ ，则再比较 $d1$ 和 $d2$ 。

答案：(1) B (2) D

考点链接 3.6.1：关系代数运算

关系代数运算符有 4 类：集合运算符、专门的关系运算符、算术比较符和逻辑运算符。根据运算符的不同，关系代数运算可分为传统的集合运算和专门的关系运算。传统的集合运算包括并、交、差和广义笛卡儿积。专门的关系运算既可以从关系的水平方向进行运算，又可以向关系的垂直方向运算，包括选择、投影、连接，以及除法。

(1) 基本运算

并、差、笛卡儿积、投影和选择是 5 种基本运算，其他运算都可以通过基本运算导出。

①并：关系 R 与 S 的并是由属于 R 或属于 S 的元组构成的集合。其形式定义如下。

$$R \cup S = \{t \mid R \hat{=} t \vee S \hat{=} t\}$$

②差：关系 R 与 S 的差是由属于 R 但不属于 S 的元组构成的集合。其形式定义如下。

$$R - S = \{t \mid R \hat{=} t \wedge t \notin S\}$$

③广义笛卡儿积：两个元数分别为 n 目和 m 目的关系 R 和 S 的广义笛卡儿积是一个 $(n+m)$ 列的元组的集合。元组的前 n 列是关系 R 的一个元组，后 m 列是关系 S 的一个元组。其形式

定义如下。

$$R \times S = \{(a_1, a_2, \dots, a_m, b_1, b_2, \dots, b_n) | (a_1, a_2, \dots, a_m) \in R, (b_1, b_2, \dots, b_n) \in S\}$$

④选择: 选择是对一个二维表按条件 F 进行横向选择, 从关系 R 中选择满足给定条件的元组。其形式定义如下。

$$\sigma_F(R) = \{t | t \in R \wedge F(t) = \text{TRUE}\}$$

其中 F 为条件, 是由常量、属性名, 以及比较运算符和逻辑运算符组成的条件表达式。

⑤投影: 投影是对一个二维表进行纵向筛选, 从关系 R 中选出若干列组成新的关系。其形式定义如下。

$$\pi_A(R) = \{t[A] | t \in R\}$$

其中 A 为 R 的属性组, $t[A]$ 表示只取元组 t 中相应 A 属性的分量。

(2) 扩展的关系运算

扩展的关系运算主要包括交、连接、除法和外连接等。

①交: 关系 R 与 S 的交是由属于 R 同时又属于 S 的元组构成的集合。记为 $R \cap S$, 显然 $R \cap S = R - (R - S)$ 。

②连接: 连接运算是从两个关系 R 和 S 的笛卡儿积中选取属性间满足一定条件的元组。笛卡儿积可以看成无条件连接。连接分为 q 连接、等值连接和自然连接 3 种。

q 连接是指从关系 R 和 S 的笛卡儿积中选取属性间满足一定条件的元组, q 是比较运算符。当 q 为 “=” 时, 称之为等值连接。

自然连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, 并去掉结果集中的重复属性。

③除: 给定关系 $R(X, Y)$ 和 $S(Y, Z)$, X 、 Y 和 Z 为属性组, $R \div S$ 应满足元组在 X 上的分量值 x 的象集 Y_x 包含关系 S 在属性组 Y 上投影的集合。

题型 7: 关系数据库 SQL 语言

命题方向: 该考点主要考查 SQL 数据库体系结构、基本组成、SQL 数据定义、数据查询、数据更新和访问控制。

【典型题 25】 (2001 年上午试题 33)

在 SQL 中, 外模式一级数据结构的基本单位是 (1)。

- (1) A. 基本表 B. 视图 C. E-R 图 D. 用户表

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。

答案: (1) B

考点链接 3.7.1: SQL 数据库体系结构

SQL 是 1974 年提出的，主要包括数据查询、数据操纵、数据定义和数据控制功能，是一种通用的、功能强大的关系数据库的标准语言。

(1) SQL 的特点

SQL 语言具有综合统一、高度非过程化、面向集合的操作方式、两种使用方式、语言简洁及易学易用等特点。

SQL 是集数据定义、数据操纵和数据控制功能于一体，语言风格统一，可独立完成数据库生命周期的所有活动。不同于非关系数据模型的数据操纵语言，SQL 是高度非过程语言。当进行数据操作时，只要指出“做什么”，无须指出“怎么做”，存储路径对用户来说是透明的，提高了数据的独立性。可以在终端键盘上输入 SQL 命令，对数据库进程操作，称之为自含式语言；同时，也可以将 SQL 语言嵌入到高级语言程序中。

(2) SQL 的功能

SQL 的功能极强，只用 9 个动词就完成了 4 大核心功能。

数据查询：SELECT。

数据定义：CREATE、DROP 和 ALTER。

数据操纵：INSERT、UPDATE 和 DELETE。

数据控制：GRANT、REVOKE。

(3) SQL 支持的三级模式结构

SQL 语言支持关系数据库的三级模式结构，外模式、模式和内模式分别对应 SQL 中的视图、基本表、存储文件。

(4) SQL 的基本组成

SQL 由数据定义语言 (DDL)、交互式数据操纵语言 (DML)、事务控制、完整性及权限管理等部分组成。

【典型题 26】 (2006 年上午试题 58)

某企业的数据库系统中有如下所示的员工关系和仓库关系，每个仓库可有多名员工，但只有一名负责人。

员工关系 (employee):

员工编号 (employeeID)	员工姓名 (name)	部门 (department)	职务 (position)
001	李蕾	仓储部	组长
002	张明	仓储部	组长
003	何苗	仓储部	保管员

仓库关系 (warehouse):

仓库编号 (warehouseID)	仓库名称 (name)	地址 (address)	负责人编号 (pricipalID)
01	仓库 A	地址 A	001
02	仓库 B	地址 B	002
03	仓库 C	地址 C	003

则创建仓库表结构的 SQL 语句为___(1)___。

- (1) A. CREATE TABLE (employeeID CHAR (2) NOTNULL UNIQUE),
name CHAR (30) NOT NULL,
address CHAR (40),
pricipalID CHAR (3));
- B. CREATE warehouse (warehouseID CHAR (2) PRIMARY KEY),
name CHAR (30) ,
address CHAR (40),
pricipalID CHAR (3));
- C. CREATE TABLE warehouse (warehouseID CHAR (2) PRIMARY KEY),
name CHAR (30) NOT NULL,
address CHAR (40),
FOREIGN KEY (pricipalID) REFERENCES employee (employeeID));
- D. CREATE TABLE warehouse (warehouseID CHAR (2),
name CHAR (30) NOT NULL,
address CHAR (40),
pricipalID CHAR (3),
PRIMARY REY (warehouseID),
FOREIGN KEY (pricipalID) REFERENCES employee (employeeID));

分析:

创建基本表的语句格式:

```
CREATE TABLE <表名> (<列名 1> <数据类型>[列级完整性约束条件]
[,<列名 2> <数据类型>[列级完整性约束条件]]...
[,<表级完整性约束条件>];
```

列级完整性约束条件有 NULL(空)、UNIQUE(取值惟一)等。表级完整性约束有 PRIMARY KEY (指定主码)和 FOREIGN KEY (指定外码)。

题目中,仓库关系表以仓库编号 warehouseID 为关键字,负责人编号 pricipalID 为外部码,这属于表级完整性约束,应放在最后部分。

答案: (1) D

【典型题 27】 (2004 年下半年上午试题 21~22)

SQL 中的数据库定义功能包括对数据库、基本表、___(1)___和___(2)___的定义。

- (1) A. 数据项 B. 表 C. 索引 D. 列
(2) A. 视图 B. 列 C. 字段 D. 元组

分析:

SQL 数据定义语言提供关系模式和视图、删除关系和视图、修改关系模式的命令。数据定义的功能包括基本表的创建、修改和删除,索引的定义和删除,视图的定义、删除和更新。

答案：(1) C (2) A

考点链接 3.7.2: SQL 数据定义

(1) 基本表的创建、修改和删除

① 创建基本表

语句格式：`CREATE TABLE <表名> (<列名 1> <数据类型> [列级完整性约束条件] [, <列名 2> <数据类型> [列级完整性约束条件]]... [, <表级完整性约束条件>];`

列级完整性约束条件有 NULL(空)、UNIQUE(取值惟一)等。表级完整性约束有 PRIMARY KEY(指定主码)和 FOREIGN KEY(指定外码)。

② 修改基本表

在表中增加新列:

`ALTER TABLE <表名> [ADD <新列名> <数据类型> [完整性约束条件]];`

从表中删除原有的某列:

`ALTER TABLE <表名> [DROP <列名> [完整性约束条件]];`

修改表中的某列:

`ALTER TABLE <表名> [MODIFY <列名> <数据类型>];`

③ 删除基本表

语句格式：`DROP TABLE <表名>`

(2) 定义和删除索引

① 建立索引

语句格式:

`CREATE [UNIQUE][CLUSTER]INDEX <索引表>
ON <表名> (<列名> [<次序>], [<列名> [<次序>]] ...);`

其中,次序可选 ASC 或 DSC,前者为升序,后者为降序,默认值为 ASC。UNIQUE 表明此索引标的每一个索引值只对应惟一的数据记录。CLUSTER 表明要建立的索引是聚集索引,意为索引项的顺序是与表中记录的物理顺序一致的索引组织。

② 删除索引

语句格式：`DROP INDEX <索引表>`

(3) 视图的定义与删除

① 创建视图

语句格式：`CREATE VIEW 视图名 (列表名)`

`AS SELECT 查询子句`

`[WITH CHECK OPTION];`

WITH CHECK OPTION 表示对 UPDATE、INSERT 和 DELETE 操作时保证更新、插入或删除的行满足视图定义中的谓词条件。

② 删除视图

语句格式：`DROP VIEW <视图名>`

【典型题 28】 (2005 年下半年上午试题 47~48)

职员关系模式为 E (Eno, Ename, Dept, Eage, Eaddr), 其中 Eno 表示职员号, Ename 表示职员名, Dept 表示职员所在部门, Eage 表示年龄, Eaddr 表示职员的家庭住址。建立“开发部”(DS 表示开发部) 职员的视图 DS_E 如下, 要求进行修改、插入操作时保证该视图只有开发部的职员。

```
CREATE VIEW DS_E
AS SELECT Eno, Ename, Dept, Eage, Eaddr
FROM
WHERE (1)
```

如下 SQL 语句可以查询开发部姓“王”职员的姓名和家庭住址。

```
Select Ename, Eaddr
From DS_E
Where (2);
```

- | | |
|-------------------|---------------------------------|
| (1) A. Dept=DS | B. Dept=DS WITH CHECK OPTION |
| C. Dept='DS' | D. Dept='DS'S WITH CHECK OPTION |
| (2) A. Ename='王%' | B. Ename Like '王%' |
| C. Ename='王*' | D. Ename Like '王*' |

分析:

在视图的创建中, 如果要对 UPDATE、INSERT 和 DELETE 操作保证更新、插入或删除的行满足视图定义中的谓词条件, 就需要使用 WITH CHECK OPTION。

在 SQL 语句中, 使用字符串需要用 ' 或 " 引用。

在计算机通常的用法中, 用 “*” 表示任意长度的字符串, 但是在 SQL 中却是 “%” 表示任意长度的字符串。在搜索子字符串时, 条件语句用 Ename Like '王%'。

答案: (1) D (2) B

【典型题 29】 (2005 年上半年上午试题 24)

学生关系模式为 S(Sno, Sname, SD, Sage), 其中, Sno 表示学生学号, Sname 表示学生姓名, SD 表示学生所在系, Sage 表示学生年龄。试将下面的 SQL 语句空缺部分补充完整, 使其可以查询计算机系学生的学号、姓名和年龄。

```
SELECT Sno, Sname, Sage
FROM S
WHERE (1)
```

- (1) A. SD=计算机 B. SD='计算机' C. 'SD'=计算机 D. 'SD=计算机'

分析:

计算机是字符型数据, 在进行查询时, 需要把“计算机”用引号括起来。

答案: (1) B

【典型题 30】 （2004 年上半年上午试题 19~20）

设有一个关系 Student（学号，姓名，系名，课程号，成绩），查询至少选修了 4 门课程的学生学号、姓名及平均成绩的 SELECT 语句应该是

```
SELECT 学号, 姓名, AVG(成绩) AS 平均成绩
FROM Student
GROUP BY _____ (1)
HAVING _____ (2)
```

- (1) A. 学号 B. 姓名 C. 系名 D. 课程号
 (2) A. COUNT(DISTINCT 学号)>3 B. COUNT(课程号)>3
 C. COUNT(DISTINCT 学号)>=3 D. COUNT(课程号)>=3

分析：

GROUP BY 子句可以将查询结果表的各行按一列或多列取值相等的原则进行分组，如果分组后还要按一定的条件对这些组进行筛选，最终只给出满足条件的组，可以使用 HAVING 语句指定筛选条件。

对于本题来说，“查询至少选修了 4 门课程的学生”则是首先要将学生按学生模式的主关键字学号分组，然后筛选对于每一个分组统计课程号数目大于 3 或大于等于 4 的分组输出。

答案：(1) A (2) B

【典型题 31】 （2003 年上半年试题 17~18）

假设供应商 S 和供应情况 SPJ 的关系模式分别为 S(Sno, Sname, Status, City)和 SPJ(Sno, Pno, Jno, Qty)。SQL 语句____(1)____不能正确地查询出“零件号 Pno 等于'P3'的供应商名 Sname”，而____(2)____是能正确查询的关系代数表达式。

- (1) A. SELECT DISTINCT Sname B. SELECT DISTINCT Sname
 FROM S FROM S
 WHERE EXISTS WHERE 0<
 (SELECT * (SELECT COUNT(*
 FROM SPJ FROM SPJ
 WHERE S.Sno=SPJ.Sno WHERE S.Sno=SPJ.Sno
 AND SPJ.Pno='P3') AND SPJ.Pno='P3')
- C. SELECT Sname D. SELECT Sname
 FROM S FROM S,SPJ
 WHERE EXISTS WHERE S.Sno=SPJ.Sno
 (SELECT * AND SPJ.Pno='P3')
 FROM SPJ GROUP BY Sname
 WHERE S.Sno=SPJ.Sno
 AND SPJ.Pno='P3')

- (2) A. $\prod_{Sname(S)} \succ \sigma_{Pno=P3}(SPJ)$

- B. $\Pi_{Sname}(S \bowtie \sigma_{Pno=P3}(SPJ))$
 C. $\Pi_{Sname}(S) - \Pi_{Sname}(\sigma_{Pno \neq P3}(S \bowtie (SPJ)))$
 D. $\Pi_{Sname}(S) - \Pi_{Sname}(S \bowtie \sigma_{Pno \neq P3}(SPJ))$

分析:

供应商 S 的关系模式为 S(Sno, Sname, Status, City), 供应情况 SPJ 的关系模式为 SPJ(Sno, Pno, Jno, Qty)。由于供应商有重名的问题, 在 SELECT 查询语句的结果中应有 DISTINCT, 用于在输出的结果中去掉重复的行; 或者是先按供应商名字分组, 然后判断供应商是否供应了 P3 零件。题目中 C 选项子查询找出了供应 P3 零件的供应商, 但是外查询结果中没有使用 DISTINCT, 取得重复元组。

与 SQL 语句对应的关系代数运算中 B 是正确的。通过选择运算找出供应 P3 的各元组, 然后通过供应商号进行自然连接, 再对供应商名进行投影。选项 C 和 D 是等价的, 如果供应商既供应 P3 零件又供应其他零件, 结果将出错。

答案: (1) C (2) B

【典型题 32】 (2000 年上午试题 6)

SQL 的 SELECT 语句中“From Q”(这里 Q 为基本表名)应理解为___(1)___。

- (1) A. Q 中的元组序号
 B. 关系 Q 的元组变量
 C. 基本表 Q 的结构定义
 D. Q 中的全部元组

分析:

SELECT FORM 是从表中选择所有的元组, 它列出的是表达式求值过程中需要扫描的关系。如果需要指定某些列属性, 可在 SELECT 与 FROM 之间说明。

答案: (1) B

考点链接 3.7.2: SQL 数据查询

(1) SELECT 查询语句

语句格式:

```
SELECT [ALL | DISTINCT]<目标列表表达式>[,<目标列表表达式>]...
      FROM<表名或视图名>[,<表名或视图名>]
      [WHERE<条件表达式>]
      [GROUP BY <列名 1>[HAVING<条件表达式>]]
      [ORDER BY <列名 2>[ASC|DESC]...]
```

查询语句中 SELECT 和 FROM 是必须的, 其余部分可选。WHERE 指定结果集中元组应满足的条件。

SELECT 子句对应的是关系代数中的投影运算, 用来列出查询结果中的属性。其输出可以是列名、表达式或集函数, DISTINCT 选项可以保证查询的结果集中不存在重复组。

FROM 子句对应的是关系代数中的笛卡儿积, 它列出的是表达式求值过程中需要扫描的关系, 即在 FROM 子句中出现多个基本表或视图时, 系统首先进行笛卡儿积操作。

WHERE 子句对应关系代数中的选择谓词。

(2) 子查询和聚集函数

子查询也称嵌套查询，是指一个 SELECT-FROM-WHERE 查询块可以嵌入另一个查询块中。在 SQL 中允许多重嵌套。

聚集函数是一个值的集合为输入，返回单个值的函数。SQL 提供了 5 个预定义集函数：平均值 AVG、最小值 MIN、最大值 MAX、求和 SUM，以及计数 COUNT。

(3) 分组查询

在 WHERE BY 子句后面加上 GROUP BY 子句可以对元组进行分组。最简单的情况是将记录分组后，也可以用 HAVING 子句来筛选它们。HAVING 只能与 GROUP BY 一起使用，一旦 GROUP BY 完成了记录分组，HAVING 子句与确定元组条件的 WHERE 子句类似。当元组含有空值时，需要注意：空值在任何聚集操作中被忽视，它对求和、求平均值和计数都没有影响。例如，COUNT(*)是某个关系中所有元组数目之和，但 COUNT(A)却是 A 属性非空的元组个数之和。NULL 值又可以在分组属性中看做一个一般的值。

(4) 更名运算

SQL 提供可为关系和属性重命名的机制，通过使用 AS 语句来实现：

old-name AS new-name

AS 子句可以出现在 SELECT 子句中，也可以出现在 FROM 子句中。

(5) 字符串操作

对于字符串进行的操作通常是使用 LIKE 的模式匹配。使用两个特殊的字符来描述模式：“%”匹配任意字符串，“_”匹配任意一个字符。模式是大小敏感的。

为了使模式中包含特殊模式字符，在 SQL 中允许使用 ESCAPE 关键词来定义转义符。转义字符紧靠着特殊字符，并放在它的前面，表示该特殊字符被当成普通字符。使用 “\” 作为转义符。

(6) 视图查询

系统执行视图查询时，通常转换成等价的对基本表的查询，然后执行查询语句。系统先从数据字典中取出该视图的定义，然后将定义中的查询语句和对该视图的查询语句结合起来，形成一个修正的查询语句。

【典型题 33】 (2006 年上半年上午试题 54)

在 SQL 语句中，删除基本表的命令是 (1)，修改表中数据的命令是 (2)。

- (1) A. DESTROY TABLE B. REMOVE TABLE
 C. DELETE TABLE D. DROP TABLE
 (2) A. REVOKE B. CHANGE C. UPDATE D. MODIFY

分析：

删除表的语句格式：DROP TABLE <表名>; DELETE 命令用于删除表中的语句。

修改表中的数据命令：UPDATE <基本表名>
 SET <列名>=<值表达式>
 [WHERE <条件表达式>]

MODIFY 用于修改基本表，REVOKE 用于访问控制中收回权限。

答案：(1) D (2) C

考点链接 3.7.3: SQL 数据更新

(1) INSERT 插入语句

SQL 的插入语句有两种格式。

语句格式:

```
INSERT INTO <基本表名>[(<列名 1>,<列名 2>...)]
```

```
VALUE(<值 1>,<值 2>...)
```

```
INSERT INTO <基本表名>[(<列名 1>,<列名 2>...)]
```

```
SELECT 子句
```

(2) UPDATE 更新语句

语句格式:

```
UPDATE <基本表名>
```

```
SET<列名 1>=<表达式 1>, <列名 2>=<表达式 2>, .....
```

```
[WHERE 条件表达式]
```

(3) DELETE 删除语句

语句格式:

```
DELETE FROM <基本表名>
```

```
[WHERE 条件表达式]
```

删除语句用于从指定表中删除满足条件的行，只能从一个关系中删除元组。

题型 8: 数据库设计

命题方向: 该考点主要考查数据库设计的特点、方法、基本步骤、数据库的实施与维护。

【典型题 34】 (2001 年上午试题 7)

E-R 模型可以转换成关系模型。当两个实体间联系是 $M : N$ 联系时,它通常可转换成 (1) 个关系模式。

- (1) A. 2 B. 3 C. $M+N$ D. $M*N$

分析:

在进行关系数据库逻辑设计时,要将概念设计阶段的基本 E-R 模型转换成关系模型。

在概念结构设计阶段得到的 E-R 图是由实体、属性和联系组成的,而关系数据库逻辑设计的结果是一组关系模式的集合。将 E-R 图转换为关系模型,实际上就是将实体、属性和联系转换成关系模式。一个实体转换为一个关系模式,实体的属性就是关系的属性,实体的码就是关系的码。

一个 1:1 联系可以转换为一个独立的关系模式,也可以与任意一端对应的关系模式合并。

一个 1:N 的联系可转换为一个关系模式,或与 N 端的关系模式合并。

一个 $M:N$ 联系转换为一个关系模式，由联系转换得到的关系模式的属性集中，包含两个发生联系的实体的码以及联系本身的属性，而关系的码为各实体码的组合。

题目中，两个实体间的联系是 $M:N$ 联系，通常可转换成 3 个关系模式。

答案：(1) B

考点链接 3.8.1：数据库设计

(1) 数据库设计的特点

数据库设计的特点包括：

- ① 数据库建设是硬件、软件和干件（技术与管理的界面）的结合。
- ② 结构（数据）设计和行为（处理）设计的结合。

(2) 数据库设计的方法

数据库设计有许多方法，常见的如下。

- ① 视图模型化及视图汇总设计法。
- ② 关系模式设计法。
- ③ 新奥尔良方法。
- ④ 基于 E-R 模型的数据库设计方法。
- ⑤ 基于 3NF 的设计方法。
- ⑥ 基于抽象语法规范的设计方法。
- ⑦ 计算机辅助设计方法。

(3) 数据库设计的基本过程

按照规范设计方法，将数据库设计分为 6 个阶段：需求分析、概念结构设计、逻辑结构设计、物理结构设计、数据库实施，以及数据库的运行和维护。

(4) 数据库设计的基本步骤

① 需求分析：通过详细调查现实世界要处理的对象，充分了解原系统的工作概况，明确用户需求，确定新系统的功能。

需求分析阶段生成的结果主要包括数据和处理两个方面。数据包括对数据字典、全系统中的数据项、数据流和数据存储的描述，处理包括对数据流图和判定表、数据字典中处理过程的描述。

② 概念结构设计

概念结构的主要特点是能真实反映现实世界，包括事物和相互之间的联系，能满足用户对数据的处理要求，是对现实世界的一个真实模型；易于理解；易于更改；易于向关系、网状及层次等各种数据模型转换。

概念结构设计的方法主要有自顶向下、自底向上、逐步扩张和混合策略。一般通过 E-R 模型来描述概念结构。

设计局部 E-R 概念模型包括确定局部结构范围、定义实体、联系定义及属性分配。即利用数据抽象机制对需求分析阶段收集到的数据进行分类、组织，形成实体、实体的属性，标识实体的码，确定实体之间的联系类型，设计分 E-R 图。

设计全局 E-R 模式包括视图集成确定公共实体集类型、局部 E-R 模式的合并及消除各种冲突。

③逻辑结构设计

逻辑设计的目的是把概念设计阶段的基本 E-R 图转换成与选用的具体机器上的 DBMS 所支持的数据模型相符合的逻辑结构。

逻辑设计分为 3 个步骤。

- I 将概念模型转换为一般的关系、网状和层次模型。
- I 将关系、网状和层次模型向特定的 DBMS 支持下的数据模型转换。
- I 对数据模型进行优化。

E-R 模型向关系模型的转换应遵循以下的原则。

- I 每个实体类型转换成一个关系模式。

一个 1:1 的联系可转换为一个关系模式，或与任意一端的关系模式合并。若独立转换为一个关系模式，那么，两端关系的码及联系的属性为该关系的属性；若与一端合并，那么将另一端的码及联系的属性合并到该端。

一个 1:N 的联系可转换为一个关系模式，或与 N 端的关系模式合并。若独立转换为一个关系模式，那么，两端关系的码及联系的属性为关系的属性，而 N 端的码为关系的码。

一个 M:N 的联系转换为一个关系模式，由联系转换得到的关系模式的属性集中，包含两个发生联系的实体的码以及联系本身的属性，而关系的码为各实体码的组合。

3 个或 3 个以上多对多的联系可转换为一个关系模式，那么，诸关系的码及联系的属性为关系的属性，而关系的码为各实体的码的组合。

- I 具有相同码的关系可以合并。

④物理结构设计

对于一个给定的逻辑数据模式选取一个最适合应用环境的物理结构的过程，称为数据库的物理设计。通常对于关系数据库物理设计的主要内容包括为关系模式选择存取方法，设计关系、索引等数据库文件的物理结构。