

# 第一章 绪论

## 考点一：数据管理技术的发展

- 1、数据管理技术的发展：人工管理阶段、文件系统阶段、数据库阶段和高阶数据库阶段。
- 2、数据库阶段数据管理的特点：
  - 采用数据模型表示复杂的数据结构
  - 有较高的数据独立性
  - 数据库系统为用户提供了方便的用户接口
  - 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。
  - 增加了系统的灵活性。
- 3、数据的含义称为数据的语义，数据和数据的语义是不可分的。数据是描述事物的符号记录称为数据。
- 4、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。
- 5、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。
- 6、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。

## 考点二：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：
  - 实体：客观存在，可以相互区别的事物称为实体。
  - 实体集：性质相同的同类实体的集合。
  - 属性：实体有很多特性，每一个特性称为属性。
  - 实体标识符（关键码或键）：能唯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：
  - 字段（数据项）：标记实体属性的命名单位称为字段或数据项。
  - 记录：字段的有序集合。
  - 文件：同一类记录的集合。
  - 关键码：能唯一标识文件汇总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象关系：

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。

一对一联系	如果实体集 E1 中每一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”
一对多联系	如果实体集 E1 中每一个实体与实体集 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，那么称为 E1 和 E2 的联系是“一对多联系”，记为“1:N”
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”

### 考点三：数据抽象的级别

1、数据模型：描述数据库的机构和定义，对现实世界的数据进行抽象。数据模型的组成要素：数据结构、数据操作和数据的完整性约束条件。

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的过程，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。

概念模型	表达用户需求观点的数据全局逻辑结构的模型
逻辑模型	表达计算机实现观点的 DB 全局逻辑结构的模型
外部模型	表达用户使用观点的 DB 局部逻辑结构的模型
内部模型	表达 DB 物理结构的模型

3、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。
- (5) 通常分为概念设计、逻辑设计（第 2 和 3 步）和物理设计三个阶段。

4、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。

5、逻辑模型的分类：层次模型、网状模型、关系模型等。

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
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
关系模型	是由若干个关系模式组成的集合

6、外部模型中视图称为视图。

7、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层次：

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。外模式有若干个记录类型组成
逻辑模式	是数据库汇总全部数据的整体逻辑结构的描述
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。一个数据库只有一个内模式。	

8、两级映像：

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
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

9、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。

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。它分为逻辑数据独立性和物理数据独立性。

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
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，那么只要对内模式的修改尽量不影响逻辑模式

**考点四：数据库管理系统 DBMS**

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。

**考点五：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。

## 第二章 关系数据库

### 考点一：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。
  - 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。
  - 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。
  - 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。
  - 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。
  - 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。
  - 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。

### 考点二：关系数据结构

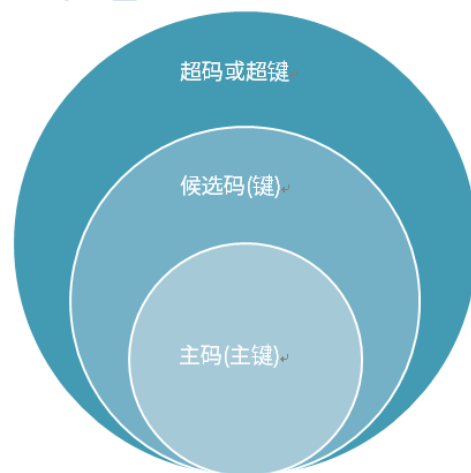
- 1、关系模型的数据结构是关系。关系有三种类型：基本关系、查询表、视图表。
  - 基本关系：也叫基本表或基表，是实际存储数据的逻辑表示。
  - 查询表：查询结果对应的表。
  - 视图表：是虚表，不对应实际存储的数据。

2、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类型。

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。关系中属性个数称为元数，元组个数称为基数。

- 关键码（简称键）：由一个或多个属性组成，包括超码、候选码、主码。
- 超码（键）：能在实体集中唯一标识一个实体的一个或多个属性的集合。
- 候选键：若关系中的某一属性或属性组的值能唯一标识一个元组，而其任意的真子集不能再标识，则称该属性组为候选码。
- 主键：一个关系中可以有多个候选码，只需选其中之一作为主码。
- 主属性：主码中的属性称为主属性。
- 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。
- 值域：关系中每一个属性都由一个取值范围，称为属性的值域。每一个属性对应一个值域，不同的属性对应于同一值域。

参照关系也成为了从关系，被参照关系也称为主关系，他们指以外码相关联的两个关系。



3、基本关系：实际存储数据的逻辑表示，关系数据库当中要求满足以下六个性质。

- 每个属性是不可分解的
- 每个关系仅仅有一种关系模式，属性的数据类型和个数固定
- 每个关系中的属性必须命名，同一关系模式中属性名不同
- 同一个关系中不允许出现候选码或候选键值完全相同的元组
- 行的顺序无所谓，可以随意交换
- 列的顺序无所谓，可以随意交换

### 考点三：关系操作集合

1、基本的关系操作包括查询操作和更新操作两大部分。

- 查询操作，关系操作中最主要部分，包括选择、投影、连接、交、差、除、并、笛卡尔积等。其中，选择、投影、并、差、笛卡尔积是 5 种基本操作。
- 更新操作，包括插入 (Insert)、删除 (Delete)、修改 (Update) 操作。

关系操作是集合操作方式，即操作的对象和结果都是集合，这种操作方式也称为一次一集合 (set-at-a-time) 方式。

2、关系代数：是一种抽象的查询语言，用对关系的运算表示查询。操作的三大要素：操作对象、操作符、操作结果。将关系代数操作分为传统的集合运算和专门的关系运算，常用的关系代数的运算符如右图所示：

运算符		含义
集合 运算符	$\cup$	并
	$-$	差
	$\cap$	交
	$\times$	笛卡尔积
专门的 关系 运算符	$\sigma$	选择
	$\Pi$	投影
	$\bowtie$	连接
	$\div$	除

(1) 传统的集合运算：包括并、差、交、笛卡尔积 4 种运算。

- 并：关系 R 与关系 S 的并运算，其结果由属于关系 R 或 S 的所有不同元组所组成。
- 差：关系 R 与关系 S 的差运算，其结果由属于关系 R 而不属于 S 的所有元组所组成。
- 交：关系 R 与关系 S 的交运算，其结果由既属于关系 R 又属于 S 的元组组成。
- 笛卡尔积：关系 R 与关系 S 的笛卡尔积运算，两个分别为 n 目和 m 目的关系 R 和 S 的笛卡尔积是一个 (m+n) 列的元组的集合。

(2) 专门的关系运算

- 选择：又称限制，他是在关系 R 中选出满足给定条件的诸元组，其中 F(t) 表示选择条件，它是一个逻辑表达式，取逻辑值“真”或“假”
- 其形式为：SELECT 关系名 WHERE 条件
- 投影：关系 R 的投影是从关系 R 中选出若干属性列组成新的关系，其形式为：PROJECTION 关系名 (属性名 1, 属性名 2, ..., 属性名 n)
- 连接：从两个关系的笛卡尔积中选取属性间满足一定关系的元组。最常用的两种连接：等值连接，自然连接。
- 除：设关系 R 除以关系 S 的结果为关系 T，则 T 包含所有在 R 但不在 S 中的属性及其值，且 T 的元组和 S 的元组的所有组合都在 R 中。

3、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。

- 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。
- 参照完整性规则：如果属性集  $K$  是关系模式  $R_1$  的主键， $K$  也是关系模式  $R_2$  的外键，那么在  $R_2$  关系中， $K$  的取值只允许两种可能，或者为空值，或者等于  $R_1$  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  $R_1$  称为参照关系， $R_2$  称为依赖关系。
- 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。

4、数据冗余：同一个数据在系统中多次重复出现。

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）

#### 考点四：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  $s$ ，都由  $t[X]=s[X]$  蕴含  $t[Y]=s[Y]$ ，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。

函数依赖根据其不同性质，可分为完全函数依赖、部分函数依赖和传递函数依赖。

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

#### 考点五：范式

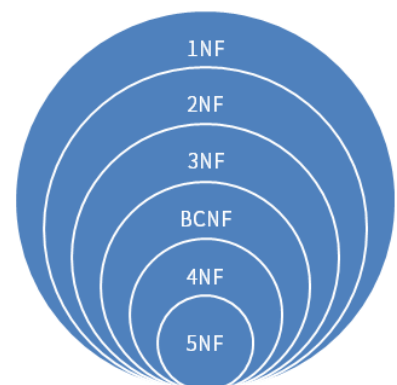
1、关系数据库中的关系需要满足一定的要求，不同程度的要求称为不同的范式 (NormalForm, NF)。

满足最低要求的称为第一范式，简称 1NF；在第一范式基础上进一步满足一些新要求的称为第二范式(2NF)；以此类推。

2、一个低一级范式的关系模式通过模式分解 (Schema decomposition) 可以转换为若干个高一级范式的关系模式的集合，这个过程叫规范化 (Normalization)

一般按属性间的依赖情况来区分关系规范化程度。

- 设  $R$  为任一给定关系，如果  $R$  中每个列与行的交点处的取值都是不可再分的基本元素，则  $R$  为第一范式。
- 若  $R \in 1NF$ ，且每一个非主属性完全函数依赖于任何一个候选码，则  $R \in 2NF$ 。



- 设  $R$  为任一给定关系，若  $R$  为 2NF，且其每一个非主属性都不传递函数依赖于候选关键字，则  $R$  为第三范式。
- 设  $R$  为任一给定关系， $X$ 、 $Y$  为其属性集， $F$  为其函数依赖集，若  $R$  为 3NF，且其  $F$  中所有函数依赖  $X \rightarrow Y$  ( $Y$  不属于  $X$ ) 中的  $X$  必包含候选关键字，则  $R$  为 BCNF。

获取完整资料 请下载 APP 或关注公众号



扫码下载 app



扫码关注公众号