

# 数据库系统原理

## 复习

@wjl

2023 年 2 月 14 日

# 目录

<b>1 重要概念</b>	<b>3</b>
1.1 数据库、数据库管理系统、数据库系统	3
1.2 三级模式和两级映像	3
1.3 物理数据独立性和逻辑数据独立性	3
1.4 数据库管理系统的作用	3
1.5 数据库管理员及其作用	3
1.6 数据模型（概念模型、结构模型）	4
1.7 数据模型的三要素：数据结构、数据操作、数据的完整性约束条件	4
1.8 关系模式	4
1.9 超码、候选码、主码	4
1.10 关系代数的主要运算	4
1.11 视图定义、视图作用	4
1.12 DDL, DML, DCL	4
1.13 数据库完整性及其主要措施	4
1.14 断言（完整性的方法，表与表之间的约束关系）	4
1.15 触发器及其功能	4
1.16 数据库安全性及其主要措施	5
1.17 范式、各范式之间的关系	5
1.18 函数依赖：完全依赖、部分依赖、传递依赖、多值依赖	5
1.19 规范化的主要目的：消除冗余	5
1.20 无损连接分解、保持依赖分解	5
1.21 索引及建立索引的优缺点	5
1.22 Hash 索引、B+ 树索引	6
1.23 事务及事务的 ACID 性质	6
1.24 并行调度	6
1.25 并行调度有可能引起的问题（丢失修改）	6
1.26 可串行化的并行调度	6
1.27 封锁	6
1.28 死锁及其解决办法	6
1.29 独占锁和共享锁、意向锁	7
1.30 两段锁协议	7
1.31 数据库常见故障类型及其恢复策略	7
<b>2 绪论</b>	<b>7</b>
2.1 概述	7
2.2 数据模型	7
2.3 数据库系统结构	8
<b>3 关系数据理论</b>	<b>9</b>
3.1 概念	9
3.2 关系模式设计中存在的两个问题	9

3.3 关系模式的码 . . . . .	9
3.4 关系模式的范式 . . . . .	9
<b>4 期中考试题</b>	<b>9</b>

# 1 重要概念

## 1.1 数据库、数据库管理系统、数据库系统

1. 数据库：数据库是长期储存在计算机内，有组织、可共享的大量数据的集合
2. 数据库管理系统：数据库管理系统是位于用户与操作系统之间的一层数据管理软件
3. 数据库系统：数据库系统是由数据库、数据库管理系统、应用程序和数据库管理员组成的存储、管理、处理和维护数据的系统

## 1.2 三级模式和两级映像

1. 模式：也称逻辑模式，是数据库中全体数据的逻辑结构和特征的描述，是所有用户的公共数据视图
2. 外模式：也称子模式或用户模式，是数据库用户能够看见和使用的局部数据的逻辑结构和特征的描述，是数据库用户的数据视图，是与某一应用有关的数据的逻辑表示
3. 内模式：也称存储模式，一个数据库只有一个内模式，是数据物理结构和存储方式的描述，是数据在数据库内部的组织方式
4. 外模式/模式映像：当模式改变时，由数据库管理员对各个外模式/模式映像作相应改变，可以使外模式保持不变。应用程序是依据数据的外模式编写的，从而应用程序不必修改，保证了数据与程序的逻辑独立性 (\*\* 数据的逻辑独立性 \*\*)
5. 模式/内模式映像：当数据库的存储结构改变时，由数据库管理员对模式/内模式映像作相应修改，可以使模式保持不变，从而应用程序也不必改变，保证了数据与程序的物理独立性 (\*\* 数据的物理独立性 \*\*)

## 1.3 物理数据独立性和逻辑数据独立性

1. 物理数据独立性：用户的应用程序与数据库中数据的物理存储是相互独立的
2. 逻辑数据独立性：用户的应用程序与数据库的逻辑结构是相互独立的

## 1.4 数据库管理系统的作用

1. 数据定义功能
2. 数据组织、存储和管理
3. 数据操纵功能
4. 数据库的事务管理和运行管理
5. 数据库的建立和维护功能

## 1.5 数据库管理员及其作用

1. 数据库管理员负责全面管理和控制数据库系统
2. 决定数据库中的信息内容和结构
3. 决定数据库的存储结构和存取策略
4. 定义数据的安全性要求和完整性约束条件
5. 监控数据库的使用和运行
6. 数据库的改进和重组、重构

## 1.6 数据模型（概念模型、结构模型）

## 1.7 数据模型的三要素：数据结构、数据操作、数据的完整性约束条件

1. 数据结构：描述数据库的组成对象以及对象之间的关系
2. 数据操作：对数据库中各种对象的实例允许执行的操作的集合
3. 数据的完整性约束条件：给定的数据模型中数据及其联系所具有的制约和依存规则

## 1.8 关系模式

1. 五元组： $R(U, D, DOM, F)$
2. F：属性组上的一组数据依赖

## 1.9 超码、候选码、主码

1. 超码 (SK)：能唯一标识关系中元组的属性或属性组
2. 候选码 (CK)：最小的超码，即其任一真子集都不是超码
3. 主码 (PK)：从候选码中选择一个最有代表性的作为主码
4. 区别：候选码是最小的超码，候选码中的属性缺一不可；超码中有多余属性。主码是候选码之一，是唯一的。一个关系可能有多个候选码，但主码是选择其中最具有代表性的

## 1.10 关系代数的主要运算

## 1.11 视图定义、视图作用

1. 视图是一个虚表，从一个或几个基本表导出，只存放视图的定义
2. 简化用户的操作
3. 使用户以多种角度看待同一数据
4. 对重构数据库提供了一定程度的逻辑独立性
5. 对机密数据提供安全保护
6. 可以更清晰地表达查询

## 1.12 DDL, DML, DCL

## 1.13 数据库完整性及其主要措施

1. 数据库完整性：数据的正确性和相容性
2. 提供定义完整性约束条件的机制
3. 提供完整性检查的方法
4. 进行违约处理

## 1.14 断言（完整性的方法，表与表之间的约束关系）

## 1.15 触发器及其功能

1. 用户定义在关系表上的一类由事件驱动的特殊过程
2. 实施更为复杂的检查和操作，具有更精细和更强大的数据控制能力

## 1.16 数据库安全性及其主要措施

1. 数据库的安全性是指指保护数据库以防止不合法使用所造成的数据泄露、更改或破坏
2. 用户身份鉴别: 即每个用户在系统中都有一个用户标识, 每个用户标识由用户名 (user name) 和用户标识号 (UID) 组成, 用户标识号在系统的整个生命周期内是唯一的。
3. 存取控制: 包括如何定义用户权限和及如何进行合法权限检查。包括自主存取控制方法和强制自主存取控制方法。
4. 视图机制、安全审计、数据库加密等

## 1.17 范式、各范式之间的关系

1. 2NF:  $R \in 1NF$ , 且每个非主属性完全函数依赖于任何一个候选码, 则  $R \in 2NF$
2. 3NF:  $R \in 1NF$ , 且每个非主属性既不传递依赖于码, 也不部分依赖于码, 则  $R \in 3NF$
3. BCNF:  $R \in 1NF$ , 每一个决定因素都含码, 则  $R \in BCNF$
- (1) 所有非主属性对每一个码都是完全函数依赖
- (2) 所有主属性对每一个不包含它的码也是完全函数依赖
- (3) 没有任何属性完全函数依赖于非码的任何一组属性
4. 4NF:  $R \in 1NF$ , 对于  $R$  的每个非平凡多值依赖,  $X$  都含有码, 则  $R \in 4NF$

1. 函数依赖: 设  $R(U)$  是属性集  $U$  上的关系模式,  $X, Y$  是  $U$  的子集, 若对于  $R(U)$  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  $Y$  上的属性值不等, 称  $X$  函数确定  $Y$ ,  $X \rightarrow Y$
2. 完全函数依赖: 在  $R(U)$  中, 如果  $X \rightarrow Y$ , 并且对  $X$  的任何一个真子集  $X'$ , 都有  $X' \not\rightarrow Y$ , 则  $Y$  对  $X$  完全函数依赖,  $X \twoheadrightarrow Y$
3. 部分函数依赖: 若  $X \rightarrow Y$ , 但  $Y$  不完全函数依赖于  $X$ , 则  $X \twoheadrightarrow Y$
4. 传递函数依赖: 在  $R(U)$  中, 如果  $X \rightarrow Y (Y \not\subseteq X)$ ,  $Y \rightarrow X$ ,  $Y \rightarrow Z (Z \not\subseteq Y)$ , 则称  $Z$  对  $X$  传递函数依赖,  $X \xrightarrow{\text{传递}} Z$
5. 多值依赖:  $R(U)$   $X, Y, Z$  是  $U$  的子集, 且  $Z = U - X - Y$ , 当且仅当对  $R(U)$  中的任一关系  $r$  给定的一对  $(x, z)$  值, 有一组  $Y$  的值, 这组值不依赖于  $x$  值而与  $z$  值无关, 多值依赖  $X \twoheadrightarrow Y$  成立。

## 1.18 函数依赖: 完全依赖、部分依赖、传递依赖、多值依赖

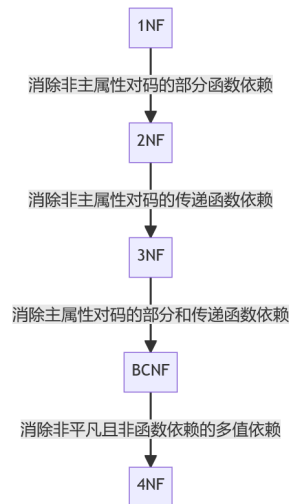
## 1.19 规范化的主要目的: 消除冗余

## 1.20 无损连接分解、保持依赖分解

1. 无损连接分解: 保证不丢失信息, 但不一定能解决数据冗余等问题
2. 保持依赖分解: 减轻或解决各种异常情况

## 1.21 索引及建立索引的优缺点

1. 加快查询速度
2. 需要占用一定的存储空间



## 1.22 Hash 索引、B+ 树索引

## 1.23 事务及事务的 ACID 性质

1. 用户定义的一个数据库操作序列，这些操作是一个不可分割的工作单位
2. 原子性、一致性、隔离性、持久性

## 1.24 并行调度

1. 单处理机系统中，事务的并行执行实际上是这些并行事务的并行操作轮流交叉运行
2. 多处理机系统中，每个处理机可以运行一个事务，实现多个事务真正的并行运行

## 1.25 并行调度有可能引起的问题（丢失修改）

1. 丢失修改：两个事务读入同一数据并修改，后一事务的修改会覆盖掉前一事务的修改，前一事务的修改丢失
2. 不可重复读：前一事务读取数据后，后一事务执行更新操作，使得前一事务无法再现前一次读取结果
3. 读脏数据：前一事务修改某一数据并写回磁盘，后一事务读取同一数据后，前一事务被撤销，被修改的数据恢复原值，后一事务读到的数据与数据库中的数据不一致

## 1.26 可串行化的并行调度

多个事务的并发执行是正确的，当且仅当其结果与按某一次序串行执行这些事务时的结果相同

## 1.27 封锁

事务对某个数据对象操作之前，先向系统发出请求，对其加锁

## 1.28 死锁及其解决办法

1. 超时法：一个事务的等待时间超过了规定的时限

2. 等待图法：有向图，反映事务的等待情况，回路说明死锁

### 1.29 独占锁和共享锁、意向锁

1. 独占锁：只允许上锁的事务读取和修改数据
2. 共享锁：上锁的事务可以读但不能修改，其他事务只能上共享锁，直到上锁的事务释放锁才可以上独占锁
3. 意向锁：对任一结点加锁时，必须先对它的上层结点加意向锁

### 1.30 两段锁协议

1. 在对任何数据进行读写操作之前，申请并获得对该数据的封锁
2. 在释放一个封锁后，事务不在申请和获得其他任何封锁

### 1.31 数据库常见故障类型及其恢复策略

1. 事务内部故障、系统故障、介质故障、计算机病毒
2. 数据转储：DBA 定期将整个数据库复制到其他存储介质上保存起来的过程
3. 登记日志文件：记录事务对数据库的更新操作的文件

## 2 绪论

### 2.1 概述

1. 基本概念
  - (1) 数据：数据是描述事物的符号记录
  - (2) 数据库：数据库是长期储存在计算机内，有组织、可共享的大量数据的集合
  - (3) 数据库管理系统：数据库管理系统是位于用户与操作系统之间的一层数据管理软件
  - (4) 数据库系统：数据库系统是由数据库、数据库管理系统、应用程序和数据库管理员组成的存储、管理、处理和维护数据的系统
2. 数据库系统的特点
  - (1) 数据结构化
  - (2) 数据的共享性高、冗余度低且易扩展
  - (3) 数据独立性高
  - (4) 数据由数据库管理系统统一管理和控制

### 2.2 数据模型

1. 组成要素
  - (1) 数据结构：描述数据库的组成对象以及对象之间的联系
  - (2) 数据操作：对数据库中各种对象的实例允许执行的操作的集合，包括操作及有关的操作规则
  - (3) 数据的完整性约束条件：一组完整性规则
2. 概念模型
  - (1) 实体：客观存在并可相互区别的事物



- (2) 属性：实体所具有的某一特性
- (3) 码：唯一标识实体的属性集
- (4) 实体型：用实体名及其属性名集合来抽象和刻画同类实体
- (5) 实体集：同一类型实体的集合
- (6) 联系：实体之间的联系通常指不同实体集之间的联系

## 2.3 数据库系统结构

### 1. 数据库系统模式的概念

模式是数据库中全体数据的逻辑结构和特征的描述

### 2. 三级模式结构

- 模式：也称逻辑模式，是数据库中全体数据的逻辑结构和特征的描述，是所有用户的公共数据视图

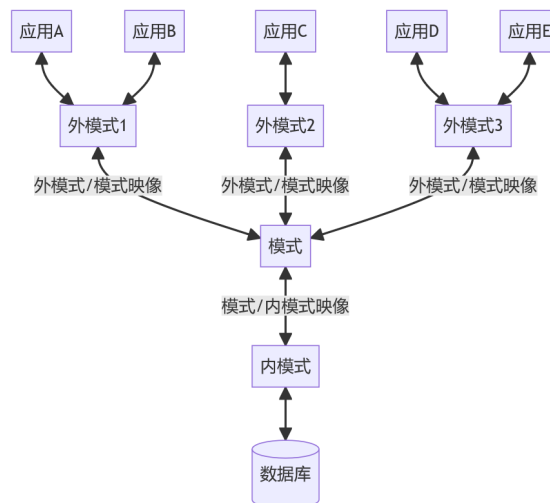
- 外模式：也称子模式或用户模式，是数据库用户能够看见和使用的局部数据的逻辑结构和特征的描述，是数据库用户的数据视图，是与某一应用有关的数据的逻辑表示

- 内模式：也称存储模式，一个数据库只有一个内模式，是数据物理结构和存储方式的描述，是数据在数据库内部的组织方式

### 3. 两级映像功能

- 外模式/模式映像：当模式改变时，由数据库管理员对各个外模式/模式映像作相应改变，可以使外模式保持不变。应用程序是依据数据的外模式编写的，从而应用程序不必修改，保证了数据与程序的逻辑独立性 (\*\* 数据的逻辑独立性 \*\*)

- 模式/内模式映像：当数据库的存储结构改变时，由数据库管理员对模式/内模式映像作相应修改，可以使模式保持不变，从而应用程序也不必改变，保证了数据与程序的物理独立性 (\*\* 数据的物理独立性 \*\*)



1. 函数依赖: 设  $R(U)$  是属性集  $U$  上的关系模式,  $X, Y$  是  $U$  的子集, 若对于  $R(U)$  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  $Y$  上的属性值不等, 称  $X$  函数确定  $Y$ ,  $X \rightarrow Y$ .
2. 完全函数依赖: 在  $R(U)$  中, 如果  $X \rightarrow Y$ , 并且对  $X$  的任何一个真子集  $X'$ , 都有  $X' \not\rightarrow Y$ , 则  $Y$  对  $X$  完全函数依赖,  $X \twoheadrightarrow Y$ .
3. 部分函数依赖: 若  $X \rightarrow Y$ , 但  $Y$  不完全函数依赖于  $X$ , 则  $X \twoheadrightarrow Y$ .
4. 传递函数依赖: 在  $R(U)$  中, 如果  $X \rightarrow Y (Y \notin X)$ ,  $Y \rightarrow X$ ,  $Y \rightarrow Z, Z \notin Y$ , 则称  $Z$  对  $X$  传递函数依赖,  $X \xrightarrow{\text{传递}} Z$ .
5. 多值依赖:  $R(U)$   $X, Y, Z$  是  $U$  的子集, 且  $Z = U - X - Y$ , 当且仅当对  $R(U)$  中的任一关系  $r$ , 给定的一对  $(x, z)$  值, 有一组  $Y$  的值, 这组值仅决定于  $x$  值而与  $z$  值无关, 多值依赖  $X \twoheadrightarrow Y$  成立.

### 3 关系数据理论

#### 3.1 概念

#### 3.2 关系模式设计中存在的两个问题

?

#### 3.3 关系模式的码

1. 候选码: 关系  $R$  的属性集  $U$  对关系中的属性或属性组  $K$  完全函数依赖, 则  $K$  为  $R$  的候选码
2. 超码: 关系  $R$  的属性集  $U$  对关系中的属性或属性组  $K$  部分函数依赖, 则  $K$  为  $R$  的超码, 候选码是最小的超码
3. 主码: 任意一个候选码

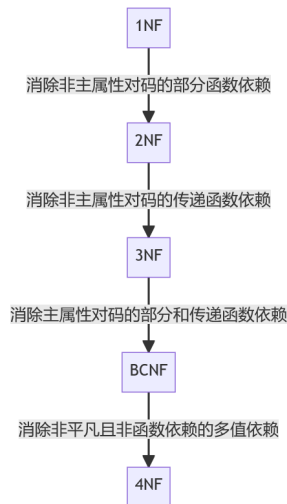
#### 3.4 关系模式的范式

1. 2NF:  $R \in 1NF$ , 且每个非主属性完全函数依赖于任何一个候选码, 则  $R \in 2NF$
2. 3NF:  $R \in 1NF$ , 且每个非主属性既不传递依赖于码, 也不部分依赖于码, 则  $R \in 3NF$
3. BCNF:  $R \in 1NF$ , 每一个决定因素都含码, 则  $R \in BCNF$ 
  - (1) 所有非主属性对每一个码都是完全函数依赖
  - (2) 所有主属性对每一个不包含它的码也是完全函数依赖
  - (3) 没有任何属性完全函数依赖于非码的任何一组属性
4. 4NF:  $R \in 1NF$ , 对于  $R$  的每个非平凡多值依赖,  $X$  都含有码, 则  $R \in 4NF$

### 4 期中考试题

1. 用数据库系统进行数据管理是由文件系统发展而来, 试简单分析文件系统与数据库系统的区别和联系。

文件系统和数据库系统的区别是: 文件系统面向某一应用程序, 共享性差, 冗余度大, 数据独立性差, 记录内有结构, 整体无结构, 由使用程序自己控制。数据库系统共享性高,



冗余度小，具有较高的物理独立性和一定的逻辑独立性，整体结构化，用数据模型描述，由数据库管理系统提供数据的安全性、完整性、并发控制和恢复能力。

文件系统和数据库系统的联系是：文件系统和数据库系统都是计算机系统中管理数据的软件。文件系统是操作系统的重要组成部分；而 DBMS 是独立于操作系统的软件。但是 DBMS 是在操作系统的基础上实现的；数据库中数据的组织和存储是通过操作系统中的文件系统来实现的。

2. 请解释什么是超码、候选码、主码，并分析它们之间的主要区别。

超码 (SK)：能唯一标识关系中元组的属性或属性组。

候选码 (CK)：最小的超码，即其任一真子集都不是超码。

主码 (PK)：从候选码中选择一个最有代表性的作为主码。

区别：候选码是最小的超码，候选码中的属性缺一不可；超码中有多余属性。主码是候选码之一，是唯一的。一个关系可能有多个候选码，但主码是选择其中最具有代表性的。

3. 试简单分析数据库管理系统 (DBMS) 的主要功能。

数据库管理系统 DBMS 是系统软件。其主要功能有：

- 数据定义：DBMS 提供数据定义语言 DDL (Data Definition Language)，用于建立、修改数据库的库结构。

- 数据操作：DBMS 提供数据操作语言 DML (Data Manipulation Language)，供用户实现对数据的追加、删除、更新、查询等操作。

- 数据库的运行管理：数据库的运行管理功能是 DBMS 的运行控制、管理功能，包括多用户环境下的并发控制、安全性检查和存取限制控制、完整性检查和执行、运行日志的组织管理、事务的管理和自动恢复，即保证事务的原子性。这些功能保证了数据库系统的正常运行。

- 数据组织、存储与管理：DBMS 要分类组织、存储和管理各种数据，包括数据字典、用户数据、存取路径等，需确定以何种文件结构和存取方式在存储级上组织这些数据，如何实现数据之间的联系。数据组织和存储的基本目标是提高存储空间利用率，选择合适的存取方法提高存取效率。

- 数据库的保护：DBMS 对数据库的保护通过 4 个方面来实现：数据库的恢复、数据库的并发控制、数据库的完整性控制、数据库安全性控制。DBMS 的其他保护功能还有系统缓冲区的管理以及数据存储的某些自适应调节机制等。

- 数据库的维护：这一部分包括数据库的数据载入、转换、转储、数据库的重组与重构以及性能监控等功能，这些功能分别由各个使用程序来完成。

4. 什么是数据独立性？试分析在数据库系统中保持数据独立性有什么益处，并分析物理数据独立性和逻辑数据独立性的区别。

数据的独立性是数据库系统的最基本的特征之一。数据独立性是指应用程序和数据结构之间相互独立，互不影响。数据独立性也是数据库系统的一个最重要的目标之一。它能使数据独立于应用程序，数据与程序的独立，把数据的定义从程序中分离出去，加上数据的存取又由 DBMS 负责，从而简化了应用程序的编制，大大减少了应用程序的维护和修改。

物理独立性是指用户的应用程序与存储在磁盘上的数据库中数据是相互独立的。即，数据在磁盘上怎样存储由 DBMS 管理，用户程序不需要了解，应用程序要处理的只是数据的逻辑结构，这样当数据的物理存储改变了，应用程序不用改变。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，即，当数据的逻辑结构改变时，用户程序也可以不变。

5. 什么是数据库安全性？试分析在数据库管理系统中实现数据库安全性控制的常用方法和技术。

数据库的安全性是指保护数据库以防止不合法使用所造成的数据泄露、更改或破坏。常用的保护措施有：

- 用户身份鉴别：即每个用户在系统中都有一个用户标识，每个用户标识由用户名 (user name) 和用户标识号 (UID) 组成，用户标识号在系统的整个生命周期内是唯一的。

- 存取控制：包括如何定义用户权限和如何进行合法权限检查。包括自主存取控制方法和强制自主存取控制方法。

- 视图机制

- 安全审计

- 数据库加密等