

课程总复习

- 数据库系统基本概念（数据模型，体系结构）
- 关系数据库
- 关系数据库标准语言SQL
- 数据库保护
- 关系数据理论
- 数据库设计
- 存储管理与索引
- 查询处理和查询优化
- 事务处理技术
- 数据库技术新发展

第一部分
基础理论

第二部分
设计理论

第三部分
实现技术

第四部分
新技术

第一章 概述

1. 数据管理技术发展的三个阶段

1) 文件系统的特点与缺点

特点：文件形式在外存中；一定独立性；文件-应用一一对应
缺点：分散、独立、冗余度高；容易数据不一致、维护难度大

2) 数据库系统数据管理的特点

数据结构化；数据冗余度小；数据、程序独立性；
统一数据控制功能；最小单位是数据项（而非记录）

2. 数据模型

1) 概念模型（信息世界）基本概念

2) 数据模型（机器世界）

三要素：
1. 数据结构
2. 数据操作
3. 完整性约束

(1) 数据模型三要素

层次：有向树，只能表示二元一对多关系，子节点必须从父节点；
网状：有向图，结构复杂，仍然是二元一对多；
关系：可直接表示多对多，关系表格、有数学基础。

(2) 数据模型分类（层次、网状、关系）及每种模型的特点

3. 数据库系统结构

1) 三级模式体系结构构成及其优点

外模式、模式、内模式

映像1 映像2

优点：保证数据独立性、数据共享和安全性、方便使用

2) 数据独立性的概念：数据物理独立性，数据逻辑独立性

3) DBMS的主要功能、DBS的组成

DBMS功能：数据定义、组织存储管理、存取、建立维护
组成：语言编译处理程序、系统运行控制程序、系统建立和维护程序、数据字典

4) 数据字典的概念

由一系列表组成，存储着数据库中有关信息的当前描述，包括数据库的三级模式、用户名表、用户权限等信息

第二章 关系数据库

1. 关系模型

1) 关系模型的数据结构

关系

2) 关系模型的完整性约束

(1) 实体完整性

主码，非空

(2) 参照完整性

外键：要么为空，要么是参照关系主码的值

(3) 用户定义完整性

3) 关系的性质

列同质、属性名不同、元组unique、行列顺序无所谓，1NF

2. 关系运算的数学基础

1) 关系运算的分类

关系代数、元组关系演算、域演算

2) 关系代数的基本运算与核心运算

基本运算：并差交 Selection Projection
核心运算：选择、投影、连接（NJ）

3) 复杂查询的关系代数表达

3. 简单元组演算的表达

第三章 关系数据库标准语言SQL

1. 关系数据语言的分类及特点

一体化、非过程化、面向集合、独立or嵌套
分类：关系代数、关系演算（元组关系演算、域关系演算）

2. SQL 数据查询功能

1) 基本查询操作

FROM A, B WHERE
ORDER BY XXX ASC/DESC WHERE ... = (SELECT ... FROM) SELECT COUNT(*)
投影、选取、排序、连表、子查询嵌套、并、差、交、库函数
、分组、算术表达式、部分匹配 UNION, MINUS, INTERSECT
GROUP BY xxx HAVING [Condition]

2) 复杂查询表达

LIKE/ NOT Like 'xx%' %_

基本查询操作的复合

3. SQL 数据定义功能

1) 表的定义

CREATE TABLE NAME
(
列名 数据类型 列级约束
表级约束
)

2) 建立视图

Create View Name
As
SELECT

3) 建立索引

Create Index Name
On TableName (ColName ASC, Col2 DESC)

4. SQL数据增、删、改功能

Insert INTO TableName (col1, col2)
Values(val1, val2)

Update Table
Set Col=xxx
Where ...

Delete From TName
Where....

第四章 数据库保护

1. 数据安全性保护

保护数据库以防止不合法的使用所造成的数据泄漏、更改和破坏。

1) 数据库安全性的含义

2) 数据库系统安全性控制的手段或机制：用户标识与认证、访问控制（存取控制），加密技术，审计、视图机制

3) 数据存取控制方法基本概念：自主存取控制，强制存取控制

4) SQL中存取控制功能

Grant xxx (On xxx) To
Revoke xxx From

授权

比较level
相等可写
>=可读

2. 数据完整性保护

1) 数据完整性的含义

正确性：合法类型、取值范围内；相容性：同一对象在不同关系数据符合逻辑

2) 完整性约束条件的分类：静态的列、元组、关系约束，动态的列、元组、关系约束

静态：不变时候，应该满足条件；动态：前后值之间应该满足约束

3) 数据库完整性控制功能构成

定义功能、检查功能、违约响应

4) SQL完整性支持功能（CREATE TABLE）

NULL/NOT NULL
UNIQUE
PRIMARY KEY

FOREIGN KEY; CHECK

第五章 关系数据理论

1. 函数依赖

- 1) 函数依赖基本概念 X 函数
- 2) 属性集闭包概念及求法 属性集闭包： X^+ ：X确定A 能通过Armstrong和F推出
- 3) 候选码求法 LRN分析，X+环路任选
- 4) 函数依赖集的最小依赖集求法 三条规则
- 5) Armstrong公理系统内容及使用 自反律（集合确定子集）、增广律（X确定Y，则XZ确定YZ）、传递律（X Y、Y Z则XZ）

2. 范式

- 1) 范式的概念及范式等级判定：**1NF, 2NF, 3NF, BCNF** 原子约束 去除：非主属性对码的部分依赖 依赖关系左侧包含码
- 2) 基于函数依赖图的模式分解方法 去除非主属性对码的传递依赖

3. 模式分解算法

- 1) 无损分解的判定准则； 判定算法（列表）；充分必要条件：R1交R2是R1或R2的一个后选码
- 2) 分解的保持函数依赖性判定方法； 简单取并集、挨个看原来X推Y，现在还能不能推出Y
- 3) 模式分解的原则 1.先求3NF函数依赖保持：Fmin，然后去掉没在F的属性，按F规则左侧分组，涉及属性被包含的删掉，自然分组。 2.基础上再判断无损性：保证一个码被包含在一个分组中。
- 4) 模式分解算法：达到3NF且满足两条分解原则的方法

第六章 数据库设计

1. 数据库规范设计法

6阶段：需求分析、概念结构设计、逻辑结构设计、物理结构、数据库实施、数据库建立和维护

1) 数据库设计分几个阶段

2) 数据库设计每个阶段的基本任务

需求：数据流图+数据字典
概念：生成概念模型
逻辑：数据模型
物理：存储结构和存取方法
实施：建立
运维：运维

2. 概念结构设计

1) E-R法

2) 基于E-R法构造局部E-R图

3) 局部E-R图综合形成总E-R图

3. 逻辑结构设计

1) E-R图向关系模型的转换规则

2) 关系模型的优化

4. 常用存取方法基本概念及选择原则

经常在查询中出现的属性、连接属性、库函数属性

1) 索引方法：B树索引，B+树索引，HASH索引

2) 聚集方法 某一属性值相同的存在一起

第七章 存储管理与索引

1. 数据存储结构/物理结构：

1) DBMS存储管理的目标

最小化传输次数

2) 文件中记录的几种组织方式的含义：堆，顺序，索引，散列，聚集

散列：hash+桶，聚集：相同存一起

随便放、按索引码顺序排列、按索引（主文件+索引表）

3) 页/块的分槽结构

Header元信息汇中：槽数组（每个元组起始偏移）、最后一个偏移量；数据从尾向前，Header从前往后

2. 缓冲区管理：缓冲区管理器的主要功能

负责缓存空间分配，内外存交换

3. 索引

1) 索引的基本概念

2) 索引的分类

(1) 两种基本类型：排序索引，哈希索引

(2) 聚集索引与非聚集索引

聚集：就是按索引域值排；非聚集：索引顺序和真正存储不同

(3) 稠密索引与稀疏索引

3) B树和B+树的原理

控制每层节点个数，使其相对平衡

4) HASH索引的原理

只比较在同一桶内的元素

第八章 查询处理与查询优化

1. 查询处理的步骤

查询分析、查询检查、查询优化、查询执行

2. 主要查询操作的实现方法

1) 选择运算

全表扫描、索引扫描

2) 连接运算

嵌套循环、索引连接、排序合并、hash join

3. 查询优化的目标

降低IO代价

4. 代数优化

1) 代数优化基本概念

利用等价关系代数表达式优化

2) 关系代数等价变换规则

先做Projection和Selection，最好二者可以一起；笛卡尔积和Selection化作join，中间结果复用

3) 代数优化的一般准则

4) 查询树的优化算法

能往树下面的就往下放

5. 物理优化

1) 物理优化基本概念

选择操作算法或存取路径

2) 物理优化的规则和方法

启发式操作算法选择：选择（是否足够小；是否有索引）和连接（是否有序；是否有索引）

第九章 事务处理技术

1. 事务

用户定义的一些列操作序列，不可分割

1) 事务的概念

原子性、一致性、隔离性、持久性

2) 事务的特性：ACID特性

3) 事务处理技术的作用

2. 数据库恢复

1) 数据库恢复的基本原理与技术

2) 故障的种类

事务内部故障、系统故障、介质故障、病毒

3) 三类故障的恢复策略：事务故障，系统故障，介质故障

UNDO
REDO (已经提交) + UNDO (还未提交) 装入副本，重做日志

4) 具有检查点的恢复技术

检查点：重新开始文件记录最近检查点；
检查点以后的任务：提交的REDO，否则UNDO

3. 并发控制

三个问题：丢失修改、读脏数据、不可重复读

数据一致性

1) 并发操作可能导致的问题

2) 并发控制的原理：基本思想与基本手段

合理调度，避免互相干扰

封锁

3) 三种封锁协议

三级封锁协议：一、写时X 二、1+读时S，读完释放 三、2的读完释放改为事务结束释放

4) 事务调度可串行性概念及两段锁协议

并行调度和某一串行的结果相同，可串行性

两段锁：1.读写前要获取锁 2.释放锁后就不再获取任意

5) 死锁预防、检测和解除的方法

一次封锁、顺序封锁

超时法、等待图 破除：代价最小

4. 多粒度封锁与意向锁的概念

1) 多粒度封锁协议

多粒度树中每个节点独立加锁

2) 意向锁基本概念

节点加意向锁，说明下层节点正在被加锁

好处：加锁时不用再检查下层的封锁，只需要检查自己和上层

IS、IX、SIX (S+IX)

第十章 数据库技术新发展

强调 场地自治性 和 自治场地之间协作性

1. 分布式数据库的定义与特点

多个分布在计算机网络上的结点组成，各个节点有独立处理能力，支持局部应用，以及通过通信支持全局应用。

2. 分布式数据库系统的模式结构

多级模式：全局外模式、全局概念模式、分片模式、分布模式、局部概念模式（后面就和原来DBMS一样）

3. 数据分片方式

按行分元组

按属性分片段

水平，但是靠其他关系判定分片条件

1) 水平分片，垂直分片，导出分片，混合分片

前面三种组合

3. 分布透明性的概念

不用知道怎么存

不用知道局部的数据模型，分布/局部概念模式 映像

1) 分片透明性，位置透明性，局部数据模型透明性

不用知道怎么分，最高

4. 分布式查询处理和优化

1) 分布式查询优化的目标

降低通信代价

2) 连接运算优化方法：半连接

右边（没有完整三角形）的，相当于先和连接条件做Projection再和左边连接
可以理解为先筛选出怎么连（让这一步传输数据量小），再真正连出数据

5. 分布式事务处理

1) 分布式事务的原子性概念

组成分布事务的子事务：要么都提交，要么都回滚

2) 分布式事务的两段提交协议

协调者1+参与者n：协调者发起准备提交，参与者回复ok or not，都ok就提交；第二步：协调者做出提交决定，发送信号，参与者回复收到，日志后开始执行。