

# 数据库考点分析

---

author: 蔡明宏, 施朱博, 焦佳宇

## 一、课本内容重点分析和工作布置思路

---

### 1.1. PPT内容重点分析（精简版）

注：对于不重要章节部分的概括，仅为笔者暂时记得的部分重点，并非复习完全给出的理解，请注意甄别

注：对于下面加粗的内容，应该是首先掌握的最容易考、考点最密集的地方。

- ch1:引言内容，对全书的概括
- ch2: relational model的介绍，介绍数据库模式、关系查询语句、关系代数等
- **ch3: 对SQL语句的基本介绍：基本SQL语句、嵌套子查询、各种散知识点 (set operation、null values、聚集函数等)**
- ch4: 对中级SQL (intermediate SQL) 的介绍：join表达式的进阶、View、事务、授权等
- ch5: 对高级SQL (Advanced SQL) 的介绍：写function函数的部分有用，JDBC和ODBC就算了
- **ch6: 使用E-R模型进行数据库设计：重点在于各种E-R图讲解，作为第7章的部分前置知识比较重要**
- **ch7: 关系数据库设计：核心在于范式分解、设计等，非常重要。**
- ch8: 复杂数据类型
- ch9: Application Development 比较侧重应用上的发展，较为次要
- ch10: 大数据 (Big Data)
- ch11: 数据分析 (Data Analytics)
- ch12: 物理存储系统 (Physical Storage Systems)
- ch13: 数据存储结构 (Data Storage Structures) : 确定IO概念很重要
- **ch14: 索引 (Indexing) , 非常重要。**
- ch15: 查询处理 (Query Processing)
- **ch17: 事务处理 (Transactions) 【部分重点】**

### 1.2. 四次作业基本涉及章节

- **ch2\*2**
- **ch3\*3**
- ch4\*1
- ch5\*1
- ch6\*1
- **ch7\*4**
- ch8\*1
- ch11\*1
- ch13\*1
- **ch14\*2**
- ch17\*1

归纳：在作业中第2、3、7、14出题均超过两道

## 1.3.如何划分刷题分工

- 思考：
  - 归纳起来，我们有**三部分核心（SQL、数据库设计、索引）**，大家考试都应掌握，所以不可完全分为一人负责一区，内心可能会有点紧张。
  - 对于零散的简单知识点部分，需要明确划分，大家更有把握
- 应对：
  - 每个人都需要重点负责**两组**刷题的汇报，让自己保持手感，但负责某一章的同学前期尽可能刷不同题；然后看其他人负责了自己没做的那个核心的题，对着做。【明天讨论下自己想领哪两个】
  - 对于零散知识点，明确分开：【ch2、ch9】【ch6、ch8】、【ch10、ch11】、【ch12、ch13】、【ch15、ch17】每人领取一个

## 1.4.如何交流汇报

为了更好地应对考试，必须在题中感悟，在题中总结，所以我们的分享最好建立在题上。如果刷题中对某些知识、某类作业通解有了较深的理解，可以集中飞书讲解。

## 二、往年数据库考试题型调研

---

### ch2.

2006A、B、C第二题 关系代数

2006C 第一题 主码、候选码、超码的区别

2020第八题 关系代数 分为是否使用聚集函数两个版本

### ch3-5 sql

2006B、C 第四题 阅读断言函数 解释其作用

2020 第一、二、三（用with）题 常规sql

2020 第六题 建表、降序排序、视图

### ch6. ER模型

2006A 第一题 给定实体集与联系画ER图，属性自拟

2006B第一题 强、弱实体集概念并 画 E-R 图说明弱实体集支付依赖于强实体集

### ch7. 关系数据库设计

2006C 第五题 关系分解的三大标准（无损、保持依赖、去冗余）

2006A 第五题 求闭包与候选码

2006A 第六题、2020第四题 BC范式分解算法

2006C 第六题 判断第四范式并进行分解

2006B 第五题 正则覆盖

2006B 第六题、2020第七题第三范式分解算法

## ch14. 索引

2006A第七题、2020第六题 解释主索引和辅助索引、辅助索引顺序查找效率高吗

2006A第八题 B+树插入

2006B第八题、2020第九题 B+树删除

2006A第九题 “理想的散列函数既是均匀的又是随机的”的理解与举例

2006A第十题 对连词选择的索引文件操作对析取选择是否适用

2006B第七题 比较密集索引和稀疏索引 理解稀疏的二级索引的意义

2006C第七题 为什么引入索引，索引是否越多越好

2006C第八题 B树索引与B+树索引比较

## ch17. 事务【基本摆明考的内容，自己背】

2006A 第四题 参照完整性的检查时机与原因，举例

2006B第九题 解释事务的ACID

2006B第十题 事务并发的意义、优点

2006C第九题 五种事务状态的状态转移图

2006C第十题 日程冲突是否可序列化

## 怪

2006A、B、C第三题 元组关系运算

## 三、书本作业梳理

---

### 3.1 布置作业知识点情况

#### 1. 第一次作业 (chapter 1~3)

- key、candidate key、primary key、foreign key
- 关系代数表达式
- SQL语句

#### 2. 第二次作业 (chapter 4~7)

- SQL语句 (计算自然连接)
- SQL 自定义函数、过程 (very important) ; 断言和触发器 (mb说触发器这块不是很重要)
- 建立E-R图, E-R图向 Relation schema 的转换, 声明constraint
- 函数依赖的定义、定理、三大推论

#### 3. 第三次作业 (chapter 7~11)

- 计算属性闭包
- BCNF 判定、分解
- 多值依赖分解4-NF
- Association Rules 数据分析, Confidence、Support等的计算

#### 4. 第四次作业 (chapter 13~17)

- 块替换、读取概念的理解, 相关替换算法效率等, 用数学式表达

- B+树建立
- 位图索引搜索过程
- transaction的状态转换、一些名词解释、特性等

## 3.2 书本课后作业部分分析

### Part 1- relational languages

1. key、relation schema、tuple、domain等相关名词概念辨析
2. 关系代数表达式
3. SQL语句的理解（基本查找、join、null的比较等）
4. 视图、授权、级联删除等概念
5. SQL自定义函数、过程

### Part 2- Database Design

1. 好的设计是什么？——无损、依赖保持、无冗余
2. 设计E-R图、E-R图向关系模式转换、标明约束
3. 函数依赖相关概念、公理、推论
4. BCNF 判断、分解、询问是否无损或保持依赖
5. 3NF 分解——涉及到正则化、无关属性、属性闭包
6. 依赖保持的验证方法
7. 多值依赖消除、转为4-NF

### Part 4 Big data analytics

1. Data analytics 中Association Rules 数据分析，Confidence、Support等的计算

### Part 5 Storage

1. 存取效率分析、算法分析
2. 主索引、次索引、dense index、sparse index
3. B+树增删改查
4. 位图查找

### Part 7 Transaction management

1. 事务的性质——一致性、原子性、持久性等等概念
2. 事务模型
3. 事务冲突（冲突等价等等）
4. 事务状态转移及相应触发条件

