**数据库系统概论复习整理**

1. **第一章 绪论**
2. **数据库的四个基本概念(重点)**

**①数据，data，描述事物的符号记录称为数据，是数据库管理的基本对象， 数据与其语义是不可分的。**

**②数据库，DB，数据库是长期存储在计算机内的、有组织的、可共享的大量 数据的集合。数据库中的数据按一定的数据模型组织、描述和存储，具有较 小的冗余度、较高的数据独立性、易扩展性和共享性。**

**数据库数据特点：永久存储、可共享、有组织**

**③数据库管理系统，DBMS，是位于用户和操作系统之间的一层数据管理软 件，和操作系统一样都是计算机的基础软件。**

**功能：数据定义、数据组织、存储管理、数据操纵、数据库的事务管理和运 行管理、数据库的创建与维护、其他功能(如不同数据库之间的互相访问和 互相操作功能)**

**④数据库系统，DBS，是由数据库、数据库管理系统、应用程序、数据库管 理员(DBA)组成的存储、管理、处理和维护数据的系统**

**特点：A. 数据结构化，数据库系统实现整体数据的结构化、这是数据库系 统和文件系统的本质区别**

1. **数据共享性高、冗余度低且易扩充**
2. **数据独立性高，包括数据的物理独立性和逻辑独立性，应用程序 和数据库中数据的物理存储以及逻辑结构相互独立**
3. **数据由数据库管理系统统一管理和控制，实现了数据安全保护、 数据完整性检查、并发控制、数据库恢复等功能**
4. **数据模型（关系模型重点，概念模型次之）**

**①分类：概念模型、逻辑模型和物理模型**

**②组成要素：**

**A.数据结构，描述数据库的组成对象和对象之间的关系**

**B.数据操作，是对数据库中各种对象和实例允许的操作的集合**

**C.数据的完整性约束，是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**

**③概念模型：A.概念与作用：也叫做信息模型，用于信息世界的建模，是现 实世界到信息世界的第一层抽象，是数据库设计人员进行数 据库设计的有力工具，也是数据库设计人员和用户之间进行 交流的语言。**

**B.涉及元素：实体、属性、码、实体型、实体集、联系**

**C.表示方法：实体-联系方法，用E-R图来描述现实世界的概念 模型，E-R方法也成为E-R模型**

**④逻辑模型(数据库邻域中常用)：层次模型、网状模型、关系模型、对象关 系数据模型、半结构化数据模型(关系模型是重点)**

**A.关系模型概念：由关系数据结构、关系操作集合、关系完整 性约束三部分组成**

**B.数据结构：每个关系的数据结构是一张规范化的二维表。相 关术语：关系、元组、属性、码、域、分量、关系模式。(要求 每一个分量必须是一个不可分的数据项)**

**C.数据操作：查询、插入、删除和更新**

**D.完整性约束：实体完整性、参照完整性、用户定义完整性**

**E.优点：关系模型建立在严格的数学概念的基础上、关系模型 概念单一，所以其数据结构简单、清晰、用户易懂易用、关系 模型存取路径对用户透明，从而具有更高的数据独立性、更好 的安全保密性、简化了程序员工作**

**F.缺点：由于存取路径对用户透明，查询效率往往不如非关系 模型。**

1. **数据库的三级模式结构（不是重点）**

**①组成：模式(逻辑模式)、外模式(子模式、用户模式)、内模式(存储模式)**

**②模式：是数据库中全体数据的逻辑结构和特征的描述**

**③外模式：是用户能看见和使用的局部数据和逻辑结构特征的描述**

**④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**

**⑤外模式/模式映像：**

**⑥模式/内模式映像：**

1. **第二章 关系数据库**

**①主码、候选码、主属性、非主属性、外码、(全码)、基本关系的六条性质**

**②关系的完整性：A.实体完整性，若属性A是基本关系R的主属性，则A不 能取空值**

**B.参照完整性：若属性F是基本关系R的外码，它与基本 关系S的主码K对应(R与S不一定是不同的关系)，则对 于R中得到每个元组在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**

**C.用户定义完整性：是针对某一具体关系数据库的约束条 件，它反映某一具体应用所涉及的数据必须满足的语义要 求**

**③关系代数(熟练手写，重点)**

1. **第三章 SQL**

**1.SQL的特点（重点）**

**①综合统一。SQL语言集数据定义语言(DDL)、数据操纵语言(DML)、数据控 制语言(DCL)于一体。**

**②高度非过程化。(用SQL语言进行数据操作，只需要提出“做什么”，而无 须指明“怎么做”，因此也无须了解存取路径。)存取路径的选择以及SQL语 句的操作过程由系统自动完成。**

**③面向集合的操作方式。SQL语言采用集合操作方式，操作对象、查找结果、 插入、删除、更新操作的对象都是元组的集合。**

**④SQL语言既是自含式语言，也是嵌入式语言。(前者指能够独立地用于联机 交互地使用方式、后者指可以嵌入到高级语言程序中)**

**⑤语言简洁、易学易用**

1. **RESTRICT和CASCADE的区别**

**①restrict：表示表的删除是有限制的。要删除的表不能被其他表的约束所引用，不能有视图和触发器等。**

**②cascade：表示表的删除没有条件限制。在删除基本表的同时，相关依赖的对象(如视图)都将被删除。**

1. **SQL语句(熟练手写，重点)**

**①模式的创建与删除**

**②表的创建与删除以及修改(重点：定义表、修改表、删除表、查询表、插入修改删除数据)**

**③视图的创建与删除（重点）**

**④索引的创建与删除以及修改**

1. **不相关子查询和相关子查询**

**①子查询的查询条件不依赖于父查询的子查询叫做不相关子查询**

**②子查询的查询条件依赖于父查询的子查询叫做相关子查询**

1. **集合查询 基于派生表的查询**

**①UNION**

**②INTERSECT**

**③EXCEPT**

**④基于派生表的查询**

1. **什么是基本表？什么是视图？二者的区别和联系？视图的优点？什么视图是可以更新的？（重点）**

**①基本表：是本身独立存在的表，SQL中一个关系就对应一个基本表**

**②视图：是从一个或几个基本表(或视图)导出的表。视图本身不独立存储于数据库中，是一个虚表。即数据库中只存放视图的定义而不存放视图的数据，这些数据仍然存储在导出视图的那些基本表中。(区别与联系)**

**③视图的优点(作用)： A.视图能简化用户的操作**

**B.视图能使用户以多种角度看待同一数据**

**C.视图对重构数据库提供了一定程度的逻辑独立性**

**D.视图能够对机密数据提供安全保护**

**E.适当地利用视图可以更清晰地表达查询**

**④视图的缺点： A.性能差，因为数据库本身并不存储视图的数据，所以 对视图的查询必须转换为对基本表的查询，降低了性能**

**B.修改限制，视图的修改是受限制的。**

**⑤可更新与不可更新：基本表的行列子集视图(从单个表导出的，只去掉了 基本表的某些行和某些列，但是保留了主码的视图叫做行列子集视图)一般 是可以更新的；如果视图中的属性来自聚集函数或者表达式，那么这个视图 一定是不可以更新的。**

1. **索引**

**①**

1. **第五章 数据库完整性**
2. **什么是数据库完整性**

**①是指数据的正确性和相容性。正确性是指数据是符合现实世界语义、反映当前实际状况的；相容性是指数据库同一对象在不同关系表中的数据是符合逻辑的。**

1. **DBMS为维护数据库的完整性而实现的功能**

**①提供定义完整性约束条件的机制。**

**②提供完整性检查的方法。一般在insert、update、delete之后执行检查**

**③进行违约处理。如果发现用户的操作违背了完整性约束条件，那么就采取一定的动作来保证数据的完整性**

2. **第六章 关系数据理论**
3. **各类码的概念**

**①超码：如果U函数依赖于K，则K称为超码**

**②候选码：**

**③主码：**

**④外码**

1. **四种范式**

**①1NF：每一个分量必须是不可分的数据项**

**②2NF：每一个非主属性完全依赖于任何一个候选码，那么关系模式就满足第二范式**

**③3NF：每一个非主属性既不传递依赖于码也不部分依赖于码，那么关系模式就满足第三范式(化为3NF的办法：把传递依赖于码的属性及其依赖的属性分离出去形成一个新表，并再原来的表中保留外键；3NF是无损的，可以保证函数依赖不丢失)**

**④BCNF：**

1. **规范化**
2. **第七章 数据库设计**
3. **第十章 数据库恢复技术**
4. **事务（重点）**

**①概念：事务是用户定义的一个数据库操作序列，这些操作要么全做要么全不做，是一个不可分割的工作单位。(一个事务可以是一条SQL语句、一组SQL语句或者整个程序，一般一个程序中包含多个事务)**

1. **事务的ACID特性（重点）**

**①原子性(A)：事务是数据库的逻辑工作单位事务中包括的各操作要么都做，要么都不做**

**②一致性(C)：事务执行的结果必须是使数据库从一个一致性状态变到另一个一致性状态。(当数据库只包含成功事务提交的结果时，就说数据库处于一致性状态)**

**③隔离性(I)：一个事务的执行不能被其他事务干扰。即一个事务的内部操作以及使用的数据对其他并发事务是隔离的。**

**④持续性(D)：也叫永久性，指一个事务一旦提交，它对数据库中数据的改变就应该是永久性的。**

**⑤ACID特性被破坏的因素：多个事务并行运行时，不同事务的操作交叉执行；事务在运行过程中被强行停止。**

1. **数据库恢复？？**
2. **第十一章 并发控制（重点）**
3. **为什么要并发控制？并发控制能保证事务的哪些特性？**

**①数据库是共享资源，通常有多个事务在运行。当多个事务并发地存取数据库时就可能产生同时读取或者修改同一数据的情况。如果不进行并发控制就可能会出现存取和存储不正确的数据。**

**②并发控制可以保证事务的一致性和隔离性。**

1. **并发操作可能会产生哪几类数据不一致？用什么方法来避免各种不一致的情况？**

**①并发操作可能会带来三类不一致:**

**A.丢失修改：两个事务t1和t2读入同一数据并修改，其中一个提交的结果破坏了另一个提交的结果，导致被破坏的修改的丢失**

**B.不可重复读：是指t1读取某一数据后，t2对这一数据进行了修改，当t1再次读该数据时无法再现前一次读取的结果。又分为三种情况：**

**C.读脏数据：**

**②避免不一致的方法就是并发控制。常用的并发控制方法有：封锁方法、时间戳方法、乐观控制法、多版本并发控制方法等**

1. **什么是封锁？基本的封锁类型有几种？**

**①封锁的概念：封锁就是事务t在对某个数据对象进行操作之前，先向系统发出请求，对其加锁。加锁后事务t对该数据对象有了一定的控制，在事务t释放它的锁之前，其他的事务不能更新或者读取此数据对象**

**②基本的封锁类型：排他锁(X锁)和共享锁(S锁)**

**A.X锁，也叫做写锁，如果事务t对数据A加上X锁，那么只允许t读 取和修改A，其他任何事务不能再对A加任何类型的锁，直到t释放A 上的锁**

**B.S锁，也叫做读锁，如果事务t对数据A加上S锁，那么事务t可以读 取A但不能修改A，其他事务只能再对A加S锁，而不能加X锁，直到 t释放A的锁为止**

1. **封锁协议（缺少严格的两段锁协议）**

**①封锁协议的概念：**

**②一级封锁协议：事务在修改数据之前必须先对数据加X锁，直到事务结束才释放。防止了丢失修改。**

**③二级封锁协议：在一级封锁协议的基础上增加事务在读取数据之前必须先对其加S锁，读完后即可释放S锁。防止了丢失修改和读脏数据。**

**④三级封锁协议：在一级封锁协议的基础上增加事务在读取数据之前必须先对其加S锁，直到事务结束才释放。防止丢失修改、读脏数据以及不可重复读的问题**

1. **活锁和死锁**

**①活锁的概念：如果事务t1封锁了数据r，事务t2请求封锁r，于是t2等待；t3也请求封锁r，当t1释放了r上的锁之后，系统首先批准了t3的请求，于是t2要继续等待，以此类推可能导致t2永远等待**

**②活锁的解决办法：采用先来先服务的策略，数据对象上的锁一旦释放就批准申请队列中的第一个事务获得锁**

**③事务t1在等待t2，而t2同时也在等待t1的情况，t1和t2两个事务永远不能结束，形成死锁。**

**④死锁的解决：？**

1. **可串行化调度和冲突可串行化调度**

**①可串行化调度：多个事务的并发执行是正确的，当且仅当其结果和某一次序串行地执行这些事务时地结果相同，称这种调度策略为可串行化调度**

**②可串行性是并发事务正确的调度准则。当且仅当它是可串行化的，才认为是正确调度**

**③冲突可串行化调度：一个调度Sc在保证冲突操作的次序不变的情况下，通过交换不冲突操作的次序得到另一个调度Sc’，如果Sc’是串行的，则称调度Sc是冲突可串行化调度。**

**④关系：冲突可串行化调度是可串行化调度的充分条件**

1. **两段锁协议**

**①概念：所有事务必须分为两个阶段对数据项加锁和解锁：第一个阶段也叫扩展阶段，在对任何数据进行读写操作之前，首先申请并获得对该数据的封锁；第二个阶段也叫做收缩阶段，在释放一个锁之后，事务不能再申请获得任何其他封锁**

1. **封锁的粒度**

1. **考试**
2. **数据库系统的基本思想与概念 DB DBMS DBS 数据模型**
3. **关系模型 基本概念（理论上讲n个集合的笛卡尔积叫做关系；由行和列组成的二维表） 要素：数据结构(关系)、数据操作（并交差补 选择 投影）、完整性约束 关系代数（重点是数据查询的代数表达！）**
4. **SQL SQL特点 基本（单表 多表 嵌套 重点是正确的逻辑表达！） 视图 索引 完整性约束和访问控制（DCL，不是重点）**

**通过JDBC访问数据库（有JDBC包 导入包 什么方法注册系统程序 通过什么方法和数据库连接）**

1. **数据库应用和系统开发--JDBC方式访问数据库有哪些步骤，还有哪些方式**
2. **数据库设计 ER模型给向关系模型映射 数据库系统的生存周期模型（清楚） 数据库设计3阶段三阶段（概念设计 逻辑设计 物理设计） 概念设计：ER 逻辑设计：ER映射为关系模型 四种复杂情况**
3. **并发控制和事务设计（幻读） 事务的概念 ACID特性解释 并发控制的必要性和基本思想 调度的概念 冲突可串行化调度的概念 关系（如果事务最终都提交并且满足两段锁协议 那就是冲突可串行化调度？/） 死锁活锁 两段锁协议和严格的两段锁协议（概念 好处 缺点 用于解决什么问题） 数据库恢复的思想，依赖于日志（immediately delay） 四种组合策略（效率最高：有数据更新立即刷新不强制等待 undo redo?） 索引有几种 索引提高查询性能的情况有哪些 不同情况下应该建立什么索引 查询按步分析？ XML MongoDB**
4. **查询性能分析？ 分析查询效率的方法（explain方法？） 如何优化？（主要是减少IO次数 寻道？旋转等待？）——一种方法：索引**
5. **非结构化数据的表示与存储、大数据时代下的NoSQL技术**
6. **填空和简答（30分）： 数据模型的三要素 关系模型特点 事务概念以及ACID特性 生存周期主要阶段 。。。**
7. **数据查询：关系代数和SQL（40分）**
8. **并发控制（10分）**
9. **数据库设计（20分）概念设计（ER图+文字说明） 概念设计的结果映射为关系模式**
10. **关系模式规范化 给出R求候选码 逐渐分解为满足BCNF要求的范式**

























