**一、单项选择题（20分，每题2分）考察基本概念，大多数人得14分以上**

1. 关于数据库的三级模式架构，下列说法正确的是( )。  
   (A) 一个数据库可以有多个内模式 (B) 视图就是内模式  
   (C) 三级模式架构实现了数据独立性 (D) 一个数据库只能有一个外模式
2. 下列关于候选键的说法，错误的是( )。  
   (A) 从一个候选键中删除任意一个属性后都不再是候选键  
   (B) 向一个候选键中添加任意其他属性后都不再是候选键  
   (C) 一个关系只有一个候选键  
   (D) 关系的主键一定是候选键
3. 已知B是关系R(A, B)的外键，参照关系S(C, D)的属性C。根据DBMS的不同配置，当执行SQL语句DELETE FROM S WHERE C = 1时，不可能发生的是( )。  
   (A) 删除R中B = 1的元组 (B) 将S中C = 1的元组的C属性值置NULL  
   (C) 拒绝执行该SQL语句 (D) 将R中B = 1的元组的B属性值置NULL
4. 在一个ER图中，实体型A和实体型B之间存在联系型R。如果对于A的任意实体，至少有1个B的实体与其建立起类型为R的联系，那么R是( )联系型。  
   (A) 1:1 (B) N:1 (C) 1:N (D) M:N
5. 已知关系模式R(A, B, C, D)上的函数依赖集为F = {AB→D, D→C}，R最高属于( )。  
   (A) 1NF (B) 2NF (C) 3NF (D) BCNF
6. 关系的文件组织形式不包括( )。  
   (A) 堆文件 (B) 有序文件 (C) 哈希文件 (D) B+树文件
7. 一个事务提交后，被它修改过的数据对象必须全部被写入磁盘上的数据文件。这个特性是事务的( )。  
   (A) 原子性 (B) 一致性 (C) 隔离性 (D) 持久性
8. 关于并发事务的调度，下列说法正确的是( )。  
   (A) 可串行化调度一定是串行调度  
   (B) 可串行化调度一定是冲突可串行化调度  
   (C) 全部冲突可串行化调度都能够通过2PL产生  
   (D) 2PL一定能够产生冲突可串行化调度
9. 如果不希望并发事务产生脏读异常，那么应将事务隔离级别最低设置为( )。  
   (A) READ UNCOMMITTED (B) READ COMMITTED  
   (C) REPEATABLE READ (D) SERIALIZABLE
10. 已知缓冲池使用NO-STEAL和FORCE策略。在进行故障恢复时，下列说法正确的是( )。  
    (A) 已提交事务必须被redo (B) 已提交事务必须被undo  
    (C) 不完整事务必须被redo (D) 不完整事务必须被undo

**二、查询题（14分）大多数人得12分以上，覆盖全部重要知识点**

已知关系数据库中包含如下关系：

* Books(ISBN, title, year, publisher)记录图书的信息，其中ISBN是书号（字符串型，主键），title是书名（字符串型），year是出版年份（日期型），publisher是出版社名（字符串型）。
* Authors(AID, name, country)记录作者信息，其中AID是作者唯一标识号（字符串型，主键），name是姓名（字符串型），country是国籍（字符串型）。
* Write(ISBN, AID, order)记录图书作者信息，其中ISBN是书号（外键，参照Books.ISBN），AID是作者唯一标识号（外键，参照Authors.AID），order是作者顺序（整型，最小值为1），该关系的主键是(ISBN, AID)。

完成下列查询：

1. （3分）将下面的SQL语句改写为关系代数表达式：   
   SELECT ISBN, title, year  
   FROM Books NATURAL JOIN Write NATURAL JOIN Authors  
   WHERE name = ‘老舍’;  
     
   简单，3个采分点，每个1分
2. （3分）将下面的关系代数表达式改写为SQL语句：

简单，3个采分点，每个1分

1. （3分）使用SQL编写查询：“哪些出版社2022年出版的图书数量超过100本？”列出出版社名和出版图书的数量。  
     
   简单，3个采分点，每个1分
2. （5分）使用SQL编写查询：“哪些出版社2022年出版的图书数量最多？”列出出版社名、书号、书名。提示：使用含有比较运算符的嵌套查询。

难度中等，4-5个采分点。可以探索新题型，例如让学生补全一个不完整的查询

**三、设计题（14分）大多数人得10分以上，覆盖全部重要知识点**

为田径比赛设计一个关系数据库，管理比赛的信息。具体需求如下：

* 记录每场比赛的信息，包括竞赛项目（例如男子100米、女子200米等）、组别（例如预赛第1组、半决赛、决赛等）、比赛时间、参赛选手号码、姓名、道次、成绩、名次、其他信息（退赛、犯规、破纪录等）。
* 记录参赛选手信息，包括号码（每个选手的号码唯一）、姓名、性别、出生日期、参赛队。
* 记录参赛队的信息，包括参赛选手、教练员姓名（可能有多个教练员）、队伍积分。

回答下列问题：

1. （10分）根据上述需求，设计数据库的ER模型，绘制出ER图，要求标明实体型、实体型属性、键属性、联系型、联系型属性、联系型的基数比、联系型的参与度。  
     
   简单，控制ER图大小，不宜耗费太长时间，采分点8个左右，含1个较难的知识点
2. （4分）将前一题设计的ER图转换为关系数据库模式。关系模式的表示方法为“关系名(属性1, 属性2, 属性3, …)”，其中主键用下划线标出。不必设计属性类型和关系的外键。  
     
   简单，4个采分点，每个1分

**四、关系数据库规范化（5分）大多数人得4分以上**

已知关系模式R(A, B, C, D, E, G)的函数依赖集为F = {AB→DE, A→D, AD→E, C→G}，回答下列问题：

1. （3分）使用Armstrong公理及其导出规则证明F逻辑蕴含ABC→EG，即从F能够推出ABC→EG。  
     
   简单，3个采分点，每个1分
2. （2分）判断关系模式属于第几范式。  
     
   简单，2个采分点，每个1分

**五、原理应用题（5分）大多数人得4分以上**  
动态展哈希表索引项重分配  
  
简单，3-4个采分点

**五、原理应用题（5分）大多数人得3分以上**

设缓冲池中只有4个页框，编号从1到4。缓冲池使用LRU（Least Recently Used）页面替换策略。缓冲池页表当前的内容如下：

|  |  |
| --- | --- |
| 页号 | 页框号 |
| 20 | 4 |
| 30 | 3 |
| 10 | 1 |
| 40 | 2 |

所有页框的状态如下：

|  |  |  |  |
| --- | --- | --- | --- |
| 页框 | Pin\_count值 | Dirty值 | 最后一次被访问的时间 （值越大表示时间越晚） |
| 1 | 1 | 0 | 111 |
| 2 | 0 | 1 | 222 |
| 3 | 0 | 1 | 444 |
| 4 | 0 | 0 | 333 |

如果现在要访问页号为50的页，那么缓冲区管理器应执行哪些操作？  
  
中等难度，3-4个采分点  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
**六、原理应用题（7分）大多数人得5分以上**

已知一个B+树中存储具有定长键值的索引项，每个节点最多存储3个键值，最少存储1个键值。该B+树的当前状态如下：



回答下列问题：

1. （3分）绘制插入键值为0013的索引项后得到的B+树。  
     
   简单，3个采分点，每个1分
2. （4分）绘制删除键值为0002的索引项后得到的B+树。  
     
   中等难度，4个采分点，每个1分

**七、原理应用题（14分）大多数人得10分以上**

已知关系R(A, B)的文件包含B(R) = 1000个页，关系S(B, C, D)的文件包含B(S) = 5000个页。R的B属性有V(R, B) = 100个不同属性值，S的B属性有V(S, B) = 250个不同属性值。另外，在S的C属性上建有二级索引。现准备执行如下SQL查询：

SELECT A, D FROM R, S WHERE R.B = S.B AND C > 60;

回答下列问题：

1. （5分）绘制解析该查询后获得的原始逻辑查询计划，并计算该计划的估计代价。  
     
   简单，5个采分点，每个1分。可以换个形式，比如比较两个给定计划的代价。
2. （3分）给出一个比原始逻辑查询计划代价更低的逻辑查询计划，并计算该计划的估计代价。  
     
   可以考察一下直方图
3. （6分）设可用缓冲区页数M = 101。对前一题你给出的逻辑查询计划，给出一个物理查询计划，并计算执行该计划所需的I/O次数。  
     
   较难，3-4个采分点

**八、原理应用题（8分）大多数人得6分以上**

已知缓冲池策略为No-Steal+No-Force。故障恢复时，WAL日志如下：

|  |  |
| --- | --- |
| 序号 | 日志记录 |
| 1 | <T1, BEGIN> |
| 2 | <T2, BEGIN> |
| 3 | <T2, X, 001, 002> |
| 4 | <T1, X, 002, 003> |
| 5 | <T1, COMMIT> |
| 6 | <BEGIN CHECKPOINT (T2)> |
| 7 | <T1, X, 003, 004> |
| 8 | <T3, BEGIN> |
| 9 | <END CHECKPOINT> |
| 10 | <T2, COMMIT> |
| 11 | <T3, X, 004, 005> |
| 12 | <T3, ABORT> |

回答下列问题：

1. （3分）对于下面4个与事务T1有关的事件，给出它们发生的时间先后顺序：日志记录<T1, BEGIN>写入磁盘，日志记录<T1, X, 003, 004>写入磁盘，数据对象X的值004写入磁盘，日志记录<T1, COMMIT>写入磁盘。  
     
   简单，3个采分点，每个1分。
2. （5分）描述基于该WAL日志的故障恢复过程。

简单，5个采分点，每个1分。

**九、（8分）大多数人得6分以上**

已知3个并发事务T1、T2和T3的一个调度如下：

|  |  |  |
| --- | --- | --- |
| T1 | T2 | T3 |
| R(X) |  |  |
|  |  | R(X) |
| R(X) |  |  |
|  | R(X) |  |
|  |  | R(Y) |
|  |  | W(Y) |
| R(Y) |  |  |
|  | W(X) |  |

回答下列问题：

1. （4分）证明该调度是冲突可串行化调度。  
     
   并发控制可以考察的知识点很多，可以考察其他知识点。
2. （4分）判断能否使用2PL协议获得上述冲突可串行化调度。如果可以，请在调度中用S-LOCK、X-LOCK和UNLOCK明确表示事务执行请求加共享锁、请求加互斥锁和解锁的动作。如果不可以，请说明原因。