

哈尔滨工业大学2020年春季学期

《数据库系统》期末考试试题（A）

注意事项

1. 考试时间：2020年7月21日8:00-10:00
2. 考试地点：线上
3. 考试形式：闭卷考试
4. 答题要求：在空白答题纸上作答，在每页答题纸上写明姓名、学号、页码
5. 交卷要求：在每页答题纸上签字，按顺序清晰拍摄全部答题纸，然后在所在考场QQ群的作业中提交全部答题纸照片原图

一、简答题（数据库语言，共20分）

已知如下关系数据库模式：

```
Student(sid, name, department, email)
Course(cid, title, credit)
Enroll(sid, cid, score)
```

- 关系Student记录学生信息，包括学号（sid）、姓名（name）、所在系（department）、电子邮件（email）。
- 关系Course记录课程信息，包括课号（cid）、课程名（title）、学分（credit）。
- 关系Enroll记录学生选课信息，包括学号（sid）、课号（cid）、成绩（score）。

回答下列问题：

1. （4分）分别用SQL和关系代数查询"Elsa"选修过的课程的课号及名称。
2. （4分）分别用SQL和关系代数查询所选课程的总学分低于120的学生的学号及所选课程的总学分。
3. （4分）分别用SQL和关系代数查询没有选修过"Database Systems"，却选修了"Data Mining"的学生的学号。
4. （2分）用SQL创建一个视图Unqualified(cid, title, cnt)，记录每门课程的不及格人数，其中cid和title的含义与数据库中的同名属性相同，cnt表示这门课不及格的学生人数。
5. （2分）是否可以在视图Unqualified上进行INSERT、DELETE或UPDATE操作？说明原因。
6. （4分）用SQL编写一个查询，说明如何根据该查询的结果验证Student的关系实例是否满足函数依赖email \rightarrow sid。

二、计算与证明题（关系数据库理论，共18分）

已知关系模式R(A, B, C, D, E, G, H)上的函数依赖集F为{ $A \rightarrow B$, $B \rightarrow D$, $AD \rightarrow EG$, $AGH \rightarrow C$ }，回答下列问题：

1. (3分) 使用Armstrong公理证明F逻辑蕴含 $A \rightarrow EG$ 。
2. (3分) 求A关于F的属性集闭包。
3. (4分) 求关系R的候选键，并说明R最高属于第几范式？
4. (4分) 求F的最小覆盖。
5. (4分) 将R分解为3个关系模式S(A, B)、T(A, E, G)和U(A, C, H)，证明该模式分解不具有函数依赖保持性。

三、论述题（共12分）

结合你的数据库系统实践经历，回答下列问题：

1. (4分) 数据独立性对于数据库系统有何重要意义？
2. (4分) 关系数据库理论对于指导关系数据库设计有哪些实际作用？
3. (4分) 事务的ACID特性中隔离性的含义是什么？DBMS是如何实际对待并发事务隔离的？

四、分析与简答题（数据库管理系统实现，共34分）

1. (4分) 在MySQL上使用下面的SQL语句创建关系t。

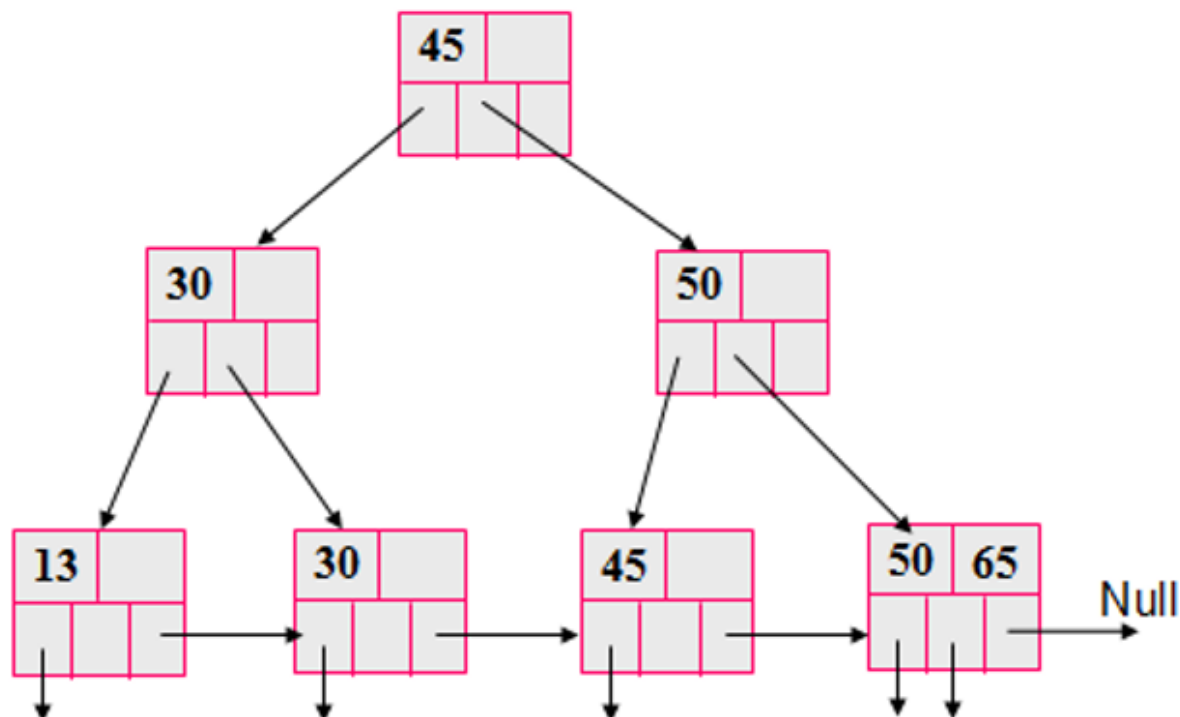
```
CREATE TABLE t (  
  id CHAR(512) PRIMARY KEY,  
  val VARCHAR(512) NOT NULL DEFAULT 'val'  
);
```

使用下面的SQL语句向t中插入元组：

```
INSERT INTO t VALUES  
  ('1', 'string1'),  
  ('2', 'string2'),  
  ... ,  
  ('3000', 'string3000');
```

如果MySQL的磁盘页面大小设置为8KB，则存储关系t大约需要100个、200个还是400个磁盘页面？请说明理由。

2. (4分) 已知下图所示的B+树，我们要从此B+树中删除键值为45的索引项，请绘制完成删除后得到的B+树。注：除根节点外，每个节点最少存储1个键值，最多存储2个键值。



3. (4分) 关系R(a, b)和S(b, c)存储在磁盘上，分别占1000和1500个磁盘块。设可用内存缓冲区块数 $M = 51$ ，R和S中的元组均未按属性b排序。回答下列问题：
- a) (2分) 证明使用排序归并连接算法执行 $R \bowtie S$ ，只需进行1趟归并。
 - b) (2分) 使用排序归并连接算法执行 $R \bowtie S$ 会产生多少次磁盘I/O？给出分析过程。
4. (6分) 设T(R)表示关系R中的元组数，V(R, A)表示关系R中A属性的不同取值个数。已知关系R(a, b)和S(b, c)的统计信息如下：
- $T(R) = 1000$, $V(R, a) = 10$, $V(R, b) = 50$
 - $T(S) = 5000$, $V(S, b) = 100$, $V(S, c) = 500$

给定如下SQL语句表示的查询：

```
SELECT c
FROM R JOIN S
WHERE a = 10;
```

回答下列问题：

- a) (2分) 绘制该查询的初始逻辑查询计划（查询表达式树），并估计该查询计划的代价。
 - b) (2分) 使用启发式查询优化方法可以对该查询进行哪些优化？绘制优化后的逻辑查询计划，并估计该查询计划的代价。
 - c) (2分) 如果用户频繁执行这种查询，每次执行时a的参数可能不同，那么在物理数据库设计阶段可以用什么方法帮助提高这种查询的执行效率？
5. (8分) 已知事务T1和T2以及如下调度：

事务T1	事务T2
R(B)	
R(A)	
W(A)	
	R(A)
	R(B)
R(C)	
W(C)	
	W(B)
	W(A)

回答下列问题：

- o a) (3分) 证明该调度是冲突可串行化调度。
 - o b) (3分) 证明使用两阶段锁协议2PL可以产生该调度。
 - o c) (2分) 尽管该调度是冲突可串行化调度，但是该调度还存在什么问题？
6. (8分) 已知DBMS使用redo/undo日志技术，WAL日志文件的全部内容如下：

```
1: <BEGIN CHECKPOINT ()>
2: <T1, START>
3: <T1, A, 1, 111>
4: <T2, START>
5: <T1, COMMIT>
6: <END CHECKPOINT>
7: <T2, B, 2, 222>
8: <T3, START>
9: ?
10: <BEGIN CHECKPOINT (T3)>
11: <T3, C, 3, 333>
```

回答下列问题：

- o a) (2分) 该日志文件的第9条日志记录被隐去了，请补全该日志记录，并说明原因。
- o b) (4分) 如果此时系统崩溃，DBMS重启，需要对哪些事务进行redo？对哪些事务进行undo？说明原因。
- o c) (2分) 故障恢复完成后，数据库对象A、B和C的值是什么？说明原因。

五、设计题（数据库设计，共16分）

注意：选择题目I和II中的一个题目作答

I. 某驾校的数据库设计需求如下：

- 管理学员信息，包括学员的身份证号、姓名、性别、出生日期、电话、所在班级。
- 管理教练信息，包括教练的身份证号、姓名、性别、出生日期、电话、入职时间。
- 管理车辆信息，包括车辆的车牌号、品牌、型号、购买日期、负责车辆维护的教练。
- 管理课程信息，包括听课班级、授课教练、授课时间、授课内容、使用车辆。
- 管理出勤信息，记录每个学员都上过哪次课。

根据需求完成下列数据库设计任务：

1. （12分）用实体-联系图（E-R图）表示该数据库的概念模型，要求明确标出以下信息：键属性（用下划线）、基数比（用1:1，1:N，M:N）、参与度（用单线或双线）。
2. （4分）将该E-R图转换为关系数据库模式，要求明确指出关系的主键和外键。

II. 任选一个你熟悉的系统，正确分析该系统的需求，绘制出相应的IDEF1X图。要求：(1) 简要叙述系统的需求--注意IDEF1X图的完备性正确性评价以此需求为基本判断依据；(2) 实体数目不得少于10个；(3) 不得出现分类联系与分类实体；(4) 将此IDEF1X图转换为对应的关系模式。

哈尔滨工业大学2020年春季学期

《数据库系统》期末考试试题（C）

注意事项

1. 考试时间：2020年7月9日18:30-20:30
2. 考试地点：线上
3. 考试形式：闭卷考试
4. 答题要求：在空白答题纸上作答，在每页答题纸上写明姓名、学号、班号、页码
5. 交卷要求：在每页答题纸上签字，按顺序清晰拍摄全部答题纸，然后在所在考场QQ群的作业中提交全部答题纸照片原图

一、简答题（数据库语言，共20分）

1. （8分）在图书管理数据库中，有如下三个关系：
 - 图书信息关系：B(B#, BNAME, AUTHOR, TYPE)，其中B#为图书编号，BNAME为书名，AUTHOR为作者，TYPE为图书类别；
 - 学生信息关系：S(S#, SNAME, CLASS)，其中S#为学号，SNAME为学生姓名，CLASS为班级号；
 - 借阅信息关系：L(S#, B#, DATE)，其中S#为借阅人学号，B#为被借阅图书编号，DATE为借阅日期。

使用关系代数回答下列问题：

- a) （2分）查询借了《红楼梦》这本书的学生的姓名。
 - b) （2分）查询105班的所有学生全都借过的图书的编号。
 - c) （2分）查询“张三”同学借过，但是“李四”同学没有借过的书的编号。
 - d) （2分）编写一个查询，说明如何根据该查询的结果验证B的关系实例是否满足函数依赖 $BNAME \rightarrow B\#$ 。
2. （12分）在博客管理数据库中，有如下三个关系：
 - 用户信息关系：U(UID#, UNAME, UAGE, USEX)，该表包含属性UID#（用户ID）、UNAME（用户姓名）、UAGE（用户年龄）、USEX（用户性别）。
 - 博客信息关系：B(BID#, UID#, BTITLE, BCONTENT, BTIME)，该表包含属性BID#（博客编号）、UID#（作者ID）、BTITLE（博客标题）、BCONTENT（博客内容）、BTIME（博客发布时间）。
 - 评论信息关系：C(CID#, BID#, UID#, CCONTENT, CTIME)，该表包含属性CID#（评论编号）、BID#（被评论博客编号）、UID#（评论者ID）、CCONTENT（评论内容）、CTIME（评论时间）。

UID #、BID #、CID#分别为关系U、B、C的主键。使用SQL语言回答下列问题：

- a) （3分）删除名为“张三”的用户发表的所有博客。
- b) （3分）查询“张三”对“李四”发表的博客的评论总数。

- o c) (3分) 定义一个视图SumU(UID#, UNAME, Year, Count#), 其中UID#为用户id, UNAME为用户姓名, Year为年份, Count#为该用户该年度发表博客的总数。
- o d) (3分) 基于视图SumU, 查询2019年发表博客数最多的用户的ID和姓名。

二、计算与简答题（关系数据库理论，共18分）

已知关系模式R(A, B, C, D, E)上的函数依赖集F为{ $AC \rightarrow E$; $B \rightarrow CE$; $C \rightarrow B$; $AB \rightarrow CD$; $AE \rightarrow D$ }, 回答下列问题:

1. (3分) 求AC关于F的属性集闭包。
2. (3分) 求R的候选键。
3. (4分) 求F的最小覆盖。
4. (4分) 将R分解为2个关系模式S(A, B, C)和T(A, B, D, E), 该分解是否是无损连接分解? 说明理由。
5. (4分) 将R进行分解一组3NF关系模式, 使得该分解既是无损连接分解, 又保持函数依赖。

三、论述题（共12分）

结合你的数据库系统实践经历, 回答下列问题:

1. (4分) 数据库的三级模式和两级映像架构对数据库系统有何重要意义?
2. (4分) 在设计大型关系数据库时, 遵照“概念设计→逻辑设计→物理设计”的设计过程有何好处?
3. (4分) 数据库管理系统的缓冲区管理与操作系统的虚拟内存管理有何不同之处?

四、分析与简答题（数据库管理系统实现，共34分）

1. (6分) 使用可扩展hash表存储整型元素。设hash函数为 $h(x) = x \bmod 16$, $h(x)$ 的结果是一个4位二进制序列, 其前d位用于确定元素x属于哪个桶。设每个桶中至多包含2个元素。现将下列元素依次添加到hash表中: 5, 19, 24, 26, 35, 48, 53, 61。初始时, hash表为空, 且 $d = 1$ 。请画出添加完上述所有元素后, 最终得到的hash表。
2. (6分) 关系R(a, b)和S(b, c)存储在磁盘上, 分别占500和1000个磁盘块。R和S上既无索引, 也未按属性b排序。设可用内存缓冲区块数为M, 回答下列问题:
 - o a) (3分) 如果 $M = 50$, 那么使用哪种连接算法执行 $R \bowtie S$ 效率最高? I/O代价是多少?
 - o b) (3分) 如果 $M = 126$, 那么使用哪种连接算法执行 $R \bowtie S$ 效率最高? I/O代价是多少?
3. (6分) 设教学管理数据库有如下3个关系模式:
 - o EMPLOYEE(ESSN#, ENAME, SALARY, DNO#)
 - o PROJECT(PNO#, PNAME)
 - o WORKS_ON(ESSN#, PNO#)
 - o DEPARTMENT(DNO#, DNAME)

用户给出以下SQL语句:

```
SELECT ENAME
FROM EMPLOYEE, WORKS_ON, PROJECT
WHERE SALARY < 3000 AND PROJECT.PNO = 'P1'
AND EMPLOYEE.ESSN# = WORKS_ON.ESSN# AND WORKS_ON.PNO# = PROJECT.PNO#;
```

回答下列问题：

- a) (3分) 写出该SQL语句对应的关系代数表达式，并画出该表达式对应的逻辑查询计划树。
- b) (3分) 使用启发式查询优化方法对以上查询计划进行优化，并画出优化后的逻辑查询计划树。

4. (8分) 已知下面两个事务：

```
T1:
read(A);
read(B);
if A > B then B := A;
write(B);

T2:
read(B);
read(A);
if B < 0 then A := B * B;
write(A);
```

回答下列问题：

- a) (3分) 简述两段锁协议（简称2PL），并给事务T1和T2添加加锁和解锁指令，使其遵从2PL协议。
- b) (3分) 按照你添加的加锁和解锁指令，给出2PL下对T1和T2的一个调度。
- c) (2分) 在什么情况下，T1和T2会发生死锁？

5. (8分) 已知DBMS使用redo/undo日志技术，WAL日志文件的全部内容如下：

```
01: <T0, START>
02: <T0, A, -30, 10>
03: <T1, START>
04: <T1, B, -20, 10>
05: <T2, START>
06: <T0, COMMIT>
07: <T2, A, 10, 20>
08: <T2, C, -20, ____>
09: <T2, COMMIT>
10: <T3, START>
11: <BEGIN CHECKPOINT ____>
12: <T3, C, 0, -20>
13: <T3, D, -30, 30>
```

回答下列问题：

- a) (2分) 该日志文件的第8条日志记录中的一个值被隐去了，请补全该日志记录，并说明原因。
- b) (2分) 该日志文件的第11条日志记录中的部分内容被隐去了，请补全该日志记录，并说明原因。
- c) (4分) 如果此时系统崩溃，DBMS重启，需要对哪些事务进行redo？对哪些事务进行undo？说明原因。

五、设计题（数据库设计，共16分）

注意：选择题目I和II中的一个题目作答

I. 某健身中心的数据库设计需求如下：

- 管理客户信息，包括客户编号、姓名、性别、出生日期、电话、身高、体重、主管教练（可以没有）。
- 管理教练信息，包括教练编号、姓名、性别、出生日期、电话、擅长的健身项目（不止一种）。
- 管理健身项目信息，包括健身项目名称、项目安排（时间、场馆、教练）。
- 管理健身记录，包括客户每次训练的开始时间、结束时间、训练场馆、教练信息、训练项目。

根据需求完成下列数据库设计任务：

1. （12分）用实体-联系图（E-R图）表示该数据库的概念模型，要求明确标出以下信息：键属性（用下划线）、基数比（用1:1，1:N，M:N）、参与度（用单线或双线）。
2. （4分）将该E-R图转换为关系数据库模式，要求明确指出关系的主键和外键。

II. 任选一个你熟悉的系统，正确分析该系统的需求，绘制出相应的IDEF1X图。要求：(1) 简要叙述系统的需求--注意IDEF1X图的完备性正确性评价以此需求为基本判断依据；(2) 实体数目不得少于10个；(3) 不得出现分类联系与分类实体；(4) 将此IDEF1X图转换为对应的关系模式。