

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。

```
SELECT cid, title
FROM Student NATURAL JOIN Course NATURAL JOIN Enroll
WHERE name = 'Elsa';

 $\Pi$  cid, title ( $\sigma$  name = 'Elsa' (Student  $\bowtie$  Course  $\bowtie$  Enroll))
```

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。

```
SELECT sid, sum(credit)
FROM Student NATURAL JOIN Enroll NATURAL JOIN Course
GROUP BY sid
HAVING sum(credit) < 120;
```

$\sigma_{\text{sum_credit} < 120}(\gamma_{\text{sid}, \text{sum(credit)} \rightarrow \text{sum_credit}}(\text{Student} \bowtie \text{Course} \bowtie \text{Enroll}))$

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。

```
SELECT sid
FROM Student NATURAL JOIN Course NATURAL JOIN Enroll
WHERE title = 'Data Mining'
EXCEPT
SELECT sid
FROM Student NATURAL JOIN Course NATURAL JOIN Enroll
WHERE title = 'Database Systems';
```

$\Pi_{\text{sid}}(\sigma_{\text{title} = \text{'Data Mining'}}(\text{Student} \bowtie \text{Course} \bowtie \text{Enroll})) - \Pi_{\text{sid}}(\sigma_{\text{title} = \text{'Database Systems'}}(\text{Student} \bowtie \text{Course} \bowtie \text{Enroll}))$

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。

```
CREATE VIEW Unqualified AS
SELECT cid, title, COUNT(*) AS cnt
FROM Student NATURAL JOIN Course NATURAL JOIN Enroll
WHERE score < 60
GROUP BY cid, title;
```

5. (2分) 是否可以在视图Unqualified上进行INSERT、DELETE或UPDATE操作? 说明原因。

不可以, 因为视图Unqualified的定义中含有GROUP BY查询。

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。

```
SELECT * FROM Student S1 JOIN Student S2 ON (S1.email = S2.email AND S1.sid != S2.sid);
```

Student的关系实例满足函数依赖email \rightarrow sid当且仅当查询结果不为空。

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 $A \rightarrow EG$ 。答: 由 $A \rightarrow B$, $B \rightarrow D$, 根据传递律, 有 $A \rightarrow D$; 再由增广律, 有 $A \rightarrow AD$; 再使用传递律, 可得 $A \rightarrow EG$ 。
2. (3分) 求A关于F的属性集闭包。答: A的闭包为{A, B, D, E, G}。
3. (4分) 求关系R的候选键, 并说明R最高属于第几范式? 答: R的候选键为AH。因为F中包含 $A \rightarrow B$, 所以存在非主属性B对候选键AH的部分函数依赖。根据定义, R属于1NF。
4. (4分) 求F的最小覆盖。答: 没有冗余的函数依赖。 $AD \rightarrow EG$ 左部含冗余属性D, $AGH \rightarrow C$ 左部含冗余属性G, 去除冗余属性后得到{ $A \rightarrow B$, $B \rightarrow D$, $A \rightarrow EG$, $AH \rightarrow C$ }。再合并左部相同的函数依赖, 得到{ $A \rightarrow BEG$, $B \rightarrow D$, $AH \rightarrow C$ }。
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。证明: 因为函数依赖 $B \rightarrow D$ 无法被三个关系上的函数依赖集合所蕴含, 故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? 答案: 当数据库的概念模式或内模式改变后, 基于外模式编写的应用程序不用改变。
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? 答: 可以帮助设计更加规范的关系数据库, 从而减少存储冗余、更新异常等问题。
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? 答: 并发事务的执行互不影响。DBMS通过4种隔离级别来实现不同程度的事物隔离。

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:

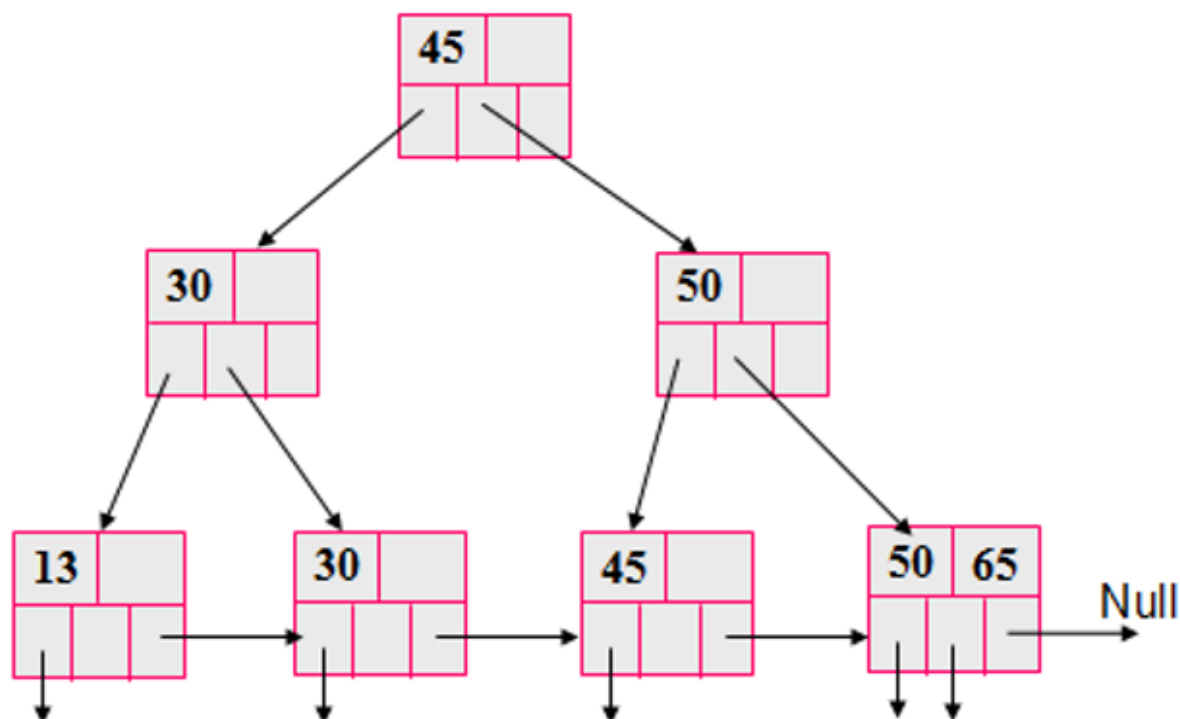
```
INSERT INTO t VALUES  
  ('1', 'string1'),  
  ('2', 'string2'),  
  ... ,  
  ('3000', 'string3000');
```

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。

答: 每个元组约占512字节, 插入3000个元组共需要大约1500000字节, 每个页面大小为8KB, 因此存储该关系实例大约需要200个页面。

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。

答：



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。答：因为 $B(R) + B(S) < M^2$ ，所以可以用1趟归并完成 $R \bowtie S$ 。
- 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。答： $3B(R) + 3B(S) = 7500$ 。

(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。答：如图所示。
 - 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。答：可将选择操作下推。如图所示。
 - 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？答：建立属性a上的索引。
4. (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：

- a) (3分) 证明该调度是冲突可串行化调度。答：等待图中只有一条从T1指向T2的边，没有环，因此是冲突可串行化调度。
- b) (3分) 证明使用两阶段锁协议2PL可以产生该调度。答：按上面文字标出的加锁和解锁顺序即可产生该调度。
- 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？答：可能存在级联中止问题。

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：

```

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：

- 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。答：<T2, COMMIT>。因为第10行的日志记录中只记录着活跃事务T3，说明T2在检查点开始前已经提交。
- 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。答：无需进行redo，需要对活跃事物T3进行undo，因为T3有可能已将部分修改写入磁盘。
- 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。答：A = 111，B =

222, C = 3。因为A和B不需要undo, C的值被undo恢复为3。

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。