

학번: 2017313107 이름: 이승태

1. 다음 문항에 대해 O, X로 답하시오. (10 점, 문항당 2 점)

- a. 표준 SQL 과 relational calculus 의 expressive power 는 동일하다. (X)
- b. RDBMS 의 모든 뷰는 갱신가능 하다. (X)
- c. unique 선언된 칼럼은 null 값을 가질 수 있다. (O)
- d. 두 개 이상의 테이블을 대상으로 정의되는 integrity constraint 기능도 있다. (O)
- e. {a,b}, {a,c}, {b,c}가 frequent itemset 이면, {a,b,c}도 frequent itemset 이다. (X)

2. 질의 작성 및 처리 관점에서, file system 기반의 host language 로 질의 로직을 구현하는 방식 대비, RDBMS 에서 SQL 질의어 및 질의최적화기가 제공하는 장점을 세 가지 열거하고 간단히 설명하시오. (각 2 점, 총 5 점)

1. SQL 로 작성하게 되면 사용자는 what I want 만 구현하면 되기 때문에 구현하기 쉽지만, file system 은 직접 code 를 다 구현해줘야 한다.
2. SQL 에서는 how 를 직접 구현해주지 않더라도 자동적으로 최적화해주기 때문에 생산성 측면에서 훨씬 이득이라고 할 수 있다. 반면 file system 으로 구현하게 되면 how 의 최적화를 직접 구현해줘야 해서 생산성이 떨어진다.
3. SQL 은 Concurrency control, crash recovery 를 지원해주지만, file system 은 원한다면 직접 구현해야 한다.

3. 수업시간에 설명한, data independence 관점에서 관계형 데이터 모델의 특징 (features)을 3 가지 나열하고 간단히 설명하시오. (5 점)

1. External schema : view 에 해당한다. Logical schema 를 기반으로 사용자에게 보여진다.
2. Logical schema: table 에 해당한다. Relations 에 해당한다.
3. Physical schema: index 에 해당하며 물리적으로 어디에 저장되어 있는지 나타낸다.

4. 보안 관점에서 뷰가 제공하는 장점을 설명하시오. 또한, 보안 이외에 뷰의 장점 2 가지를 기술하시오. (5 점).

View 는 user 에게 보여주고 싶은 것만 select, from, where 를 통해 보여줄 수 있으며, content based 하고, flexible security policy 를 제공해준다.

5. 교재에서 사용되는 스키마 Sailors(sid, sname, rating), Reserve(sid, bid, date), Boats(bid, bname, color)에서, 1) 적어도 하나의 붉은 색 배를 예약한 선원의 ID, 이름과 예약 건수, 2) 모든 붉은색(RED) 배를 예약한 선원의 이름을 구하는 SQL 질의를 작성하시오. 단, 2)의 경우 EXISTS 및 EXCEPT (or MINUS) 구문을 사용하여야 함. (10 점)

<p>1) (5 점)</p> <pre>select s.sid, s.sname, count(*) from Sailors s, Reserve r, Boats b where s.sid = r.sid and r.bid = b.bid and b.color = "RED" group by s.sid, s.sname;</pre>	<p>2) (5 점)</p> <pre>select s.name from sailors s where not exists(s</pre>
--	--

6. 오라클의 경우, primary key(PK)에 대해 인덱스를 자동생성하는데, 이 PK 인덱스는 PK 칼럼에 대한 selection 조건이 주어졌을 때 해당 질의를 빨리 처리하기 위해 사용될 수 있다. 이 이외에 entity integrity 및 referential integrity 관점에서 PK 인덱스 역할에 대해 설명하시오. (총 5 점)

1. Entity integrity 관점에서 보았을 때 만약 현재 table 에 insert 가 실행되었을 때, 빠르게 primary key 가 동일한 것이 존재하는지 체크하기 위하여, PK 인덱스를 사용한다.
2. Referential integrity 관점에서 보았을 때 다른 테이블이 이 테이블의 primary key 를 참조하고 있다면 다른 테이블에서 tuple 이 추가되었을 때, 빠르게 foreign key 가 이 테이블의 primary key 에 있는지 체크하기 위해 PK 인덱스를 사용한다.

7. 다음의 EMP, DEPT 테이블에 대해, 다음 질문에 답하시오.(총 25 점)

EMP	EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
	7839	KING	PRESIDENT		17-NOV-81	5000		10
	7698	BLAKE	MANAGER	7839	01-MAY-81	2850		30
	7782	CLARK	MANAGER	7839	09-JUN-81	2450		10
	7566	JONES	MANAGER	7839	02-APR-81	2975		20
	7654	MARTIN	SALESMAN	7698	28-SEP-81	1250	1400	30
	7499	ALLEN	SALESMAN	7698	20-FEB-81	1600	300	30
	7844	TURNER	SALESMAN	7698	08-SEP-81	1500	0	30
	7900	JAMES	CLERK	7698	03-DEC-81	950		30
	7521	WARD	SALESMAN	7698	22-FEB-81	1250	500	30
	7902	FORD	ANALYST	7566	03-DEC-81	3000		20
	7369	SMITH	CLERK	7902	17-DEC-80	800		20
	7788	SCOTT	ANALYST	7566	09-DEC-82	3000		20
	7876	ADAMS	CLERK	7788	12-JAN-83	1100		20
	7934	MILLER	CLERK	7782	23-JAN-82	1300		10

- a. 다음과 같은 deptno, job 별 cross-tab 에 필요한 measure 값들 (즉, 해당직원수, 예: 6, 14) 을 구하는 단일 SQL 을 작성하시오. (Cross-tab 형태로 결과 display 하는 질의 아님) (5 점)

정답: select deptno, job, count(*) from emp group by cube(deptno, job);				
Count	DeptNo			
Job	10	20	30	Subtotal
ANALYST		2		2
CLERK	1	2	1	4
MANAGER	1	1	1	3
PRESIDENT	1			1
SALESMAN			4	4
Subtotal	3	5	6	14

- b. 다음 질의의 의미를 설명하고, 같은 의미의 window function 질의를 작성하시오. (총 10 점)

```
SELECT e1.deptno, e1.ename, e1.hiredate, e1.sal, avg(e2.sal) avgsal
FROM emp e1, emp e2
WHERE e1.hiredate >= e2.hiredate and e1.deptno=e2.deptno
GROUP BY e1.deptno, e1.ename, e1.hiredate, e1.sal
ORDER BY deptno
```

의미 (5 점):

자신이 속한 부서의 사람들 중에서 자신보다 hiredate 가 높은 사람들의 평균 sal 을 구해서 column 에 추가(avgsal) 하고 deptno 순으로 정렬하는 쿼리이다.

Window function 질의 (5 점):

```
select e1.deptno, e1.ename, e1.hiredate, e1.sal, avg(e1.sal) over(partition by e1.deptno order by hiredate desc
range between bounded preceding and current row) as avgsal
from emp.e1;
```

(계속)

EMP

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
7839	KING	PRESIDENT		17-NOV-81	5000		10
7698	BLAKE	MANAGER	7839	01-MAY-81	2850		30
7782	CLARK	MANAGER	7839	09-JUN-81	2450		10
7566	JONES	MANAGER	7839	02-APR-81	2975		20
7654	MARTIN	SALESMAN	7698	28-SEP-81	1250	1400	30
7499	ALLEN	SALESMAN	7698	20-FEB-81	1600	300	30
7844	TURNER	SALESMAN	7698	08-SEP-81	1500	0	30
7900	JAMES	CLERK	7698	03-DEC-81	950		30
7521	WARD	SALESMAN	7698	22-FEB-81	1250	500	30
7902	FORD	ANALYST	7566	03-DEC-81	3000		20
7369	SMITH	CLERK	7902	17-DEC-80	800		20
7788	SCOTT	ANALYST	7566	09-DEC-82	3000		20
7876	ADAMS	CLERK	7788	12-JAN-83	1100		20
7934	MILLER	CLERK	7782	23-JAN-82	1300		10

- c. Oracle connect by 기능을 이용해서, 모든 직원에 대해, 그 직원의 이름과 직간접 상사 이름 쌍을 모두 출력. **힌트:** SELECT 절에서 CONNECT_BY_ROOT 키워드는 출발점 튜플을 지칭 (예: CONNECT_BY_ROOT ename 은 출발 튜플 ename 을 지칭). (5 점) 또한, SQL script 에서 배운 with subquery factoring 기능을 이용해서 같은 질의를 작성하시오(5 점).

--	--

d. 다음 두 질의 결과는 동일한지? 그리고, 그 이유를 설명하시오. (5 점)

```
SELECT EMPNO  
FROM EMP  
WHERE EMPNO NOT IN (SELECT MGR FROM EMP);
```

```
SELECT EMPNO  
FROM EMP E1  
WHERE NOT EXISTS (SELECT * FROM EMP E2 WHERE E2.MGR =E1.EMPNO);
```

결과:

다르다.

이유:

MGR 값에 null 값이 존재하기 때문이다.

MGR 이 null 값인 king 은 Not in 에서 unknown 이 되므로 출력이되므로 출력이 되지 않는다.

8. 아래 EMP 테이블에 정의된 세 가지 무결성제약조건 (primary key, foreign key, check)을 emp table insertion, update, deletion 시 구동하는 3 개의 trigger 를 이용해서 enforce 하고자 할 때, 각 trigger 의 Condition-Action 을 간단히 기술하시오. 단, PK 에 대한 update 는 허용하지 않는 것으로 가정. (총 15 점)

```
create table EMP (ename char(20) not null,  manager char(20), sal number,  
                primary key ename  
                foreign key (manager) references EMP on delete cascade,  
                check ( 200 <= sal <= 1000))
```

ANSWER: 각 5 점

1) Emp table insertion Trigger 1:

Insertion 을 진행할 때 trigger 의 Condition 은 primary key 인 ename 의 값이 겹치는 것이 없어야 한다. Sal 의 값은 200 이상 1000 이하여야 하며, manager 은 EMP(자신)를 참조하고 있으므로 ename 에 manager 의 값이 존재해야 한다. 이 때 이 조건을 만족하지 못하면 reject 되는 action 을 취한다.

2) Emp table update trigger (5 점)

Update 를 진행할 때 trigger 의 conditions 은 Sal 의 값은 200 이상 1000 이하여야 하며, manager 은 EMP(자신)를 참조하고 있으므로 ename 에 manager 의 값이 존재해야 한다. 이 때 이 조건을 만족하지 못하면 reject 되는 action 을 취한다.

3) Emp table deletion trigger (5 점)

Deletion 을 진행할 때 trigger 의 condition 은 자신의 primary key 를 참조하는 tuple 이 존재한다면, action 은 자신의 primary key 를 참조하고 있는 모든 tuple 을 지우게 된다.

9. Association Rule 마이닝 기법의 대표적인 알고리즘인 A-Priori 에 대해 논하시오. 구체적으로, frequent itemset 을 구하는 단계에서, 1) Naïve 방식의 문제점을 논하고, 2) A-priori Property 를 설명하고, 3) 그에 기반한 두 가지 최적화 기회를 설명하시오. (총 10 점).

- 1) Naïve 방식에서는 frequent itemset 을 구할 때 모든 경우를 적용하여 계산한다. 만약 item 의 개수가 N 개라고 하면, 2^N 번 만큼 database 를 체크해줘야 하므로 시간이 매우 오래 걸린다.
- 2) A-priori Property 는 만약 a,b,c 라는 item 들이 있고, {a}, {b}라는 set 이 frequent 하고, {c}라는 set 은 frequent 하지 않을 때, 이것들을 subset 으로 가지는 {a, b}는 frequent 할 가능성이 있고 {a, c}나 {b, c}의 set 은 frequent 하기 불가능하다는 알고리즘으로 위의 Naive 방식에서 보다 경우의 수를 줄여나갈 수 있다.
- 3) A-priori Property 의 최적화는 모든 subset 들이 frequent 할 때, 그 set 이 frequent 할 가능성이 있다는 것을 이용한다. 예를 들어 a, b, c, d 라는 item 이 있고, 아이টে을 2 개 선택했을 때, {a,b}, {a,c}라는 set 이 frequent 하다고 했을 때, {a,b,c}는 {b,c}가 frequent 하지 않으므로 frequent 할 가능성이 없게 된다. 이로써 경우의 수를 더 줄일 수 있게 된다.

10. 다음 transaction table, DB,에 대해 frequency 가 3 회 이상인 1 개 item 으로 구성된 frequent itemset 을 구하는 SQL 문을 작성하시오. (5 점)

DB

TID	ITEM
100	1
100	3
100	4
200	2
200	3
200	5
300	1
300	2
300	3
300	5
400	2
400	5

정답:

```
select cast(item_set as fi_num) as item_set
from DB(dbms_frequent_itemset(fi_transactional)(
  cursor(select TID for ITEM),
  (3/12),
  1,
  1,
  null,
  null
));
```

=====

===== 끝: 수고

했습니다.

=====

=====