COMP9315 DBMS Implementation
Final Exam 14s2

[Instructions] [Notes] [PostgreSQL] [C]
[Q1] [Q2] [Q3] **[Q4]** [Q5] [Q6] [Q7] [Q8]

Question 4 (10 marks)

Consider a database called "uni", similar to the one in the previous question, with the following schema:

```
People(id,title,family,given,address,gender,birthday,country)
Courses(id,code,title,uoc,convenor)
Enrolments(course,student,mark,grade)
Items(id,course,name,maxMark)
Assessments(item,student,mark)
```

Primary keys are underlined. Foreign keys are in italic. Most foreign keys use the name of the table to which they refer; `convenor` and `student` both refer to the `People` table. The `Enrolments` table links `People` to `Courses`, and records their final mark/grade for the course. The `Items` table indicates what assessment items (e.g. assignments) are in each course. The `Assessments` table tells what mark a student received for each assessment item they did.

a. Consider the following query execution plans produced by PostgreSQL for the above database:

```
uni=# explain analyze select * from Courses where id=1234;

                                  QUERY PLAN
------------------------------------------------------------------------------
 Index Scan using courses_pkey on courses
                              (cost=0.28..8.29 rows=1 width=67)
                              (actual time= 0.093..0.093 rows=0 loops=1)
   Index Cond: (id = 1234)
 Total runtime: 0.130 ms
(3 rows)

uni=# explain analyze select * from Courses where code='COMP3311';

                                  QUERY PLAN
------------------------------------------------------------------------------
 Seq Scan on courses  (cost=0.00..21.25 rows=5 width=67)
                      (actual time=0.084..0.362 rows=1 loops=1)
   Filter: (code = 'COMP3311'::bpchar)
   Rows Removed by Filter: 979
 Total runtime: 0.396 ms
```

Based on the above, answer the following:

i. Is there a `Courses` tuple with `id` 1234?
ii. What is the total number of `Courses` tuples?
iii. What is the difference in how the two queries are evaluated?
iv. Which query is the more efficient?

b. Consider the following query execution plan (slightly-edited to make it more readable):

```
uni=# explain analyze
uni-# select c.code, count(*)
uni-# from   courses c join items i on (c.id = i.course)
uni-# group  by c.code order by count(*) desc;

                                  QUERY PLAN
------------------------------------------------------------------------------
 Sort  (cost=179.71..180.21 rows=200 width=12)
       (actual time=28.368..29.481 rows=980 loops=1)
   Sort Key: (count(*))
   Sort Method: quicksort  Memory: 55kB
   ->  HashAggregate  (cost=170.07..172.07 rows=200 width=12)
                      (actual time=25.358..26.893 rows=980 loops=1)
       ->  Hash Join  (cost=31.05..150.41 rows=3931 width=12)
```

```
                    (actual time=2.916..18.444 rows=3931 loops=1)
          Hash Cond: (i.course = c.id)
       ->  Seq Scan on items i  (cost=0.00..65.31 rows=3931 width=4)
                                (actual time= 0.017..4.835 rows=3931 loops=1)
       ->  Hash  (cost=18.80..18.80 rows=980 width=16)
                 (actual time=2.882..2.882 rows=980 loops=1)
           ->  Seq Scan on courses c
                             (cost=0.00..18.80 rows=980 width=16)
                             (actual time=0.012..1.392 rows=980 loops=1)
Total runtime: 30.649 ms
```

Based on the above, answer the following:

i. Explain in english what this query is trying to do?
ii. Does the query use external merge sort for its sorting?

c. Consider the following query execution plan (slightly-edited to make it more readable):

```
                                  QUERY PLAN
------------------------------------------------------------------------------
 Nested Loop  (cost=21.59..89.19 rows=1 width=52)
              (actual time=1.541..2.354 rows=3 loops=1)
   ->  Hash Join  (cost=21.31..84.22 rows=1 width=14)
                  (actual time=1.514..2.287 rows=3 loops=1)
         Hash Cond: (e.course = c.id)
       ->  Seq Scan on enrolments e  (cost=0.00..62.83 rows=18 width=18)
                                     (actual time=0.025..1.576 rows=279 loops=1)
             Filter: (grade = 'FL'::bpchar)
       ->  Hash  (cost=21.25..21.25 rows=5 width=4)
                 (actual time=0.356..0.356 rows=1 loops=1)
           ->  Seq Scan on courses c  (cost=0.00..21.25 rows=5 width=4)
                                      (actual time=0.336..0.349 rows=1 loops=1)
                 Filter: (code = 'SOMA1641'::bpchar)
   ->  Index Scan using people_pkey on people p
                                      (cost=0.28..4.96 rows=1 width=42)
                                      (actual time=0.011..0.013 rows=1 loops=3)
         Index Cond: (id = e.student)
 Total runtime: 2.429 ms
```

Based on the above, answer the following:

i. How many students satisfied the conditions in the query?
ii. Give either an SQL statement or relational algebra expression that might have produced this execution plan.
The final result includes the fields: `Student.id`, `Student.family`, `Enrolments.mark`, `Enrolments.grade`

Instructions:

- Type your answer to this question into the file called `q4.txt`
- Submit via: **`submit q4`**

End of Question