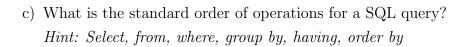# 1 Short Answer

a) When talking about distributing our databases across multiple servers, we discuss data needs as being either transactional or analytical. For both of these cases, do we prefer partitioning or duplication and why?

b) In datalog, what is "fixed point"? To what sorts of rules does this semantic apply?

c) What is the standard order of operations for a SQL query?

   Hint: Select, from, where, group by, having, order by

d) Give an example of a query that can be expressed with datalog, but not with set-semantic Relational Algebra

e) Explain the difference between OPEN and CLOSED types in semi-structured data

f) Provide two unique relational algebra expressions which provide the same final relation. Use `A,B,C...` as your starting relations as needed.

2 SQL

For this question, use a relational database storing student course records with the following schema. Use `AS` when aliasing. Primary keys are underlined.

`Student(`<u>`sid`</u>`,dept,gradyear,fname,lname)`

`Course(`<u>`dept,coursenum`</u>`,name,description)`

`Takes(sid,dept,coursenum,quarter)`

a) Write the `CREATE TABLE` statement for the table `Takes`. `Sid` is an `INT`, `dept` is a `VARCHAR(5)`, `coursenum` is an `INT` and `quarter` is a `VARCHAR(4)`.

   `Sid, dept` and `coursenum` are all foreign keys.

b) Write a query which finds the unique first and last names of all students that took UWBW 599 (Underwater Basket Weaving Dissertation)

c) Write a query to find all the students who took only classes 400-level or lower.

d) Is this query monotone? Why or why not?

3 Datalog

For the following question, use a schema around published authors and scholarly articles.

`Author(autid,fname,lname)`

`Article(articleid,name,year)`

`Authored(autid,articleid)`

Write a datalog rule degreeSeparation(autid,num) which calculates the degree of separation the author has from `Author(_,Paul,Erdos)`. You may assume there is exactly one author named Paul Erdos.

For those unfamiliar with the problem, degrees of separation indicates how close two authors are. If you imagine authors as vertices in a graph and edges representing whether those two authors have co-authored a paper, the degree of separation for two authors is the length of the *minimum* path between them.

As an aside: because this problem is difficult to describe, it is too difficult for the actual exam, but I think it illustrates the type of datalog program you should be able to write. Use as many rules as necessary.

4 Relational Algebra

Use the following schema.

Student(<u>sid</u>,dept,gradyear,fname,lname)

Course(<u>dept</u>,<u>coursenum</u>,name,description)

Takes(sid,dept,coursenum,quarter)

Write the bag-semantic relational algebra expression for the following query. You may use S for Student and T for takes without renaming. You may draw the query tree if you'd like, but your final solution should be the RA expression

```
SELECT S.sid,
       COUNT(T.coursenum) as count
FROM Student AS S
     JOIN Takes AS T
         ON S.sid = T.sid
WHERE T.coursenum < 400
GROUP BY S.sid
HAVING COUNT(T.coursenum) > 10
```