

# **CS4.301: Data and Applications (Monsoon 2022)**

## **Quiz - 3**

**Time:** 45 minutes

**Maximum Marks:** 20

Q1. Let us consider Twitter-like setting, with the following specifications.

- Users post 'tweets', that is short pieces of text.
- They may tag their tweets with zero or more tags of their own choice. For example, a user tweeting about Apple's *'Far Out'* event may decide to use the tag 'FarOut' (prefixed by a 'sharp' sign: #FarOut, if we follow the convention by the twitter site).
- A user 'u' may follow zero or more other users, which means that their 'tweets' are visible to user 'u' when he/she logs in.

For the above setting, we will use the following schema:

- Person (uname, city, street) - Assume the uname is unique.
- Follows (uname1, uname2) - Person uname1 follows person uname2
- Tweets (tid, ttext) - Tweet with tid has text ttext
- PersonTweets (uname, tid, ts) - Person uname posted tweet tid at timestamp ts
- TweetTag (tid, tagname) - Tweet tid had tagname in its list of tags.

We now want to extract some information from the database. For each of the following questions, give the corresponding SQL query.

- Find all the people (uname) who posted a tweet with tag 'MODI'. [2 marks]
- Find all the different, distinct tags ever used. [2 marks]

- iii. Find all the tags 'Rahul Gandhi' uses in his tweets. (i.e., Rahul Gandhi's tweeting interests) [2 marks]
  
  
  
  
  
  
  
  
  
  
- iv. Find all the people (uname) whose reading and tweeting interests do not intersect, i.e., who do not read the tags they write about. [4 marks]
  
  
  
  
  
  
  
  
  
  
- v. **Bonus:** Find all pairs of people (uname) who have at least one follower in common. [2 marks]

Q2. Consider a relation R(K, A, B, FK1, FK2) that satisfies all integrity constraints, where K is the key of R, foreign keys FK1 and FK2 refer to primary key K, the domain of attributes A and B are integer values {1, 2, 3, 4, 5} or NULL.

For the following parts, **select all options that are correct.**

**Part A [2 marks]**

Which of the following queries can give more rows in the result compared to the following query result?

“SELECT \* FROM R WHERE K=123;”

Note: 123 is a value taken from the domain of values of the key attribute K.

- a. SELECT DISTINCT R2.K, R2.A, R2.B, R2.FK1, R2.FK2 FROM R AS R1, R2;
- b. SELECT \* FROM R WHERE K IS NULL;
- c. SELECT \* FROM R WHERE FK1 IS NULL and FK2 IS NULL;
- d. SELECT DISTINCT R1.K, R1.A, R1.B, R1.FK1, R1.FK2 FROM R AS R1, R2 WHERE R1.A=R2.A;

**Part B [2 marks]**

Which of the following queries give the same result?

Note: At least two queries give the same result.

- a. SELECT \*  
FROM R  
WHERE A=1 AND  
K IN (SELECT K FROM R  
WHERE B=1 AND FK1 IN (SELECT FK1  
FROM R WHERE K IS NULL  
));
- b. SELECT \* FROM R WHERE FK1 IS NULL;
- c. SELECT \* FROM R WHERE A=NULL AND A=2;
- d. SELECT \* FROM R AS R1, R2 WHERE R1.K=R2.K AND R2.K<R1.K AND A=NULL;

Q3. Consider the following relations P(X, Y, Z), Q(X, Y, T) and R(Y, V)

P			Q			R	
X	Y	Z	X	Y	T	Y	V
X1	Y1	Z1	X2	Y1	3	Y1	V1
X1	Y1	Z2	X1	Y2	5	Y3	V2
X2	Y2	Z2	X1	Y1	6	Y2	V3
X2	Y4	Z4	X3	Y3	7	Y2	V2

How many tuples will be returned by the following relational algebra query? Show working. **[4 marks]**

$$\pi_X \left( \sigma_{(P.Y = R.Y \wedge R.V = V2)} (P \times R) \right) - \pi_X \left( \sigma_{(Q.Y = R.Y \wedge Q.T < 7)} (Q \times R) \right)$$

