

## Question 1 (50 points)

For this problem we consider a movie reviewing database with following three tables:

1. `Movies(title, imdb number, year)`, where the primary key is `imdb number`;
2. `Reviewers(userid, username)`, where the primary key is `userid`;
3. `Reviews(userid, imdbnumber, rating, comment)`, where the primary key is `(userid, imdb number)`, `userid` is foreign key referencing `Reviewers`, and `imdb number` is foreign key referencing `Movies`.

For these tables we know the following statistics:

- `Movies` consists of  $N_1 = 60,000$  tuples, there are:
  - 40,000 distinct movie titles,
  - 90 distinct years, 1925–2014 inclusive.
- `Reviewers` consists of  $N_2 = 50,000$  tuples, there are:
  - 50,000 distinct userids,
  - 50,000 distinct usernames
- `Reviews` consists of  $N_3 = 1,300,000$  tuples, there are:
  - 38,000 distinct userids,
  - 33,000 distinct movie titles,
  - 5 distinct ratings (i.e. 1, 2, 3, 4, and 5) without nulls.

For the queries below, assume that there are no correlations between the columns of a table nor any prior knowledge about the data (i.e., assume uniform distribution). Estimate the number of resulting tuples for the query, and give the answer with **fourth** significant digit accuracy. We will accept either rounding half up or down, but no partial credit will be given.

- (a) **SELECT \* FROM Movies WHERE year = 2001 OR year = 1924;**
- (b) **SELECT \* FROM Movies WHERE year = 1999 OR title = “Fight Club”**
- (c) **SELECT \* FROM Reviews WHERE rating > 1**
- (d) **SELECT title, count(\*) FROM Movies GROUP BY title**
- (e) **SELECT count(\*)**  
**FROM Movies JOIN Reviews ON Movies.imdb number = Reviews.imdb number**  
**GROUP BY Movies.title**

## Question 2 (50 points)

2.1 Consider the following legal instance of a relational schema  $S$  with attributes  $ABC$ :

S	A	B	C
	$\alpha$	9	T
	$\alpha$	16	F
	$\beta$	20	F

Table 1: Legal instance of schema  $S$  for question 2.1

- (a) Which of the following dependencies are *violated* by the instances of  $S$  in Table 1? Answer Yes/NO and give reason
- $A \rightarrow B$  is violated.
  - $B \rightarrow A$  is violated.
  - $C \rightarrow A$  is violated.
  - $AC \rightarrow B$  is violated.
  - $B \rightarrow AC$  is violated.
- (b) By only observing the instance of  $S$  in Table 1, can you identify the functional dependencies that hold on schema  $S$ ? YES/NO and Also give reason.

2.2 For the next set of questions consider the relational schema  $r = \{P, Q, R, S, T, U, V\}$  and the set of functional dependencies FD:

$$P \rightarrow Q \quad (1)$$

$$Q \rightarrow R \quad (2)$$

$$PS \rightarrow TRV \quad (3)$$

$$QT \rightarrow UR \quad (4)$$

$$S \rightarrow V \quad (5)$$

- (a) Which of the following is a minimum cover of the FD?
- The given FD is a minimum cover.
  - $\{P \rightarrow Q, Q \rightarrow R, PS \rightarrow T, QT \rightarrow U, S \rightarrow V\}$
  - $\{P \rightarrow Q, Q \rightarrow R, P \rightarrow T, Q \rightarrow U, S \rightarrow V\}$
  - $\{P \rightarrow Q, Q \rightarrow R, PS \rightarrow T, QT \rightarrow UR, S \rightarrow V\}$
  - none of the above - the cover is \_\_\_\_\_
- (b) Yes/No: Which of the following functional dependencies can be deduced, from the above set of functional dependencies (Eq. (1)-(5))?
- $P \rightarrow R$
  - $PS \rightarrow U$
  - $QS \rightarrow U$
  - $QST \rightarrow P$
- (c) True or False: The attribute closure  $\{Q\}^+$  is  $\{Q, R, T\}$ . Also compute it step by step
- (d) True or False: The attribute closure  $\{PS\}^+$  is  $\{P, Q, R, S, T, U, V\}$ . Also compute it step by step.