CSED 516 / Data 516: Homework 2
Supplementary Questions

Dan Suciu

October, 2019

Name: _____

Turn in your answers here `https://forms.gle/Uw1wsrtMLgrco49w7`

1. (10points)

   Consider a database schema with three relations and their sizes:

```
Likes(drinker, beer)     // 100,000 tuples
Frequents(drinker, bar)  // 2,000,000 tuples
Serves(bar, beer)        // 5,000,000 tuples
```

   Consider the following two queries:

```
Q1:
select *
from Likes, Frequents, Serves
where Likes.drinker = Frequents.drinker
  and Frequents.bar = Serves.bar

Q2:
select *
from Likes, Frequents, Serves
where Likes.drinker = Frequents.drinker
  and Frequents.bar = Serves.bar
  and Likes.beer = Serves.beer
```

   Answer the questions below:

   (a) (5 points) What is the largest possible output of Q1?

   (b) (5 points) What is the largest possible output of Q2?

2. (10points)

   Consider again Q2, and assume we compute it on a cluster with $p = 8,000$ servers. We want to compute Q2 in a single communication round.

   (a) (5 points) Compute the optimal (minimal) load per server.

   (b) (5 points) Find the optimal shares for the hypercube algorithms to compute the query Q2 in one round with the optimal load that you found at the previous step.