

## Assignment 2

CIS 452/552 Database Issues (Big Data), Winter 2017

due 11:59 pm, Friday, February 10th

1. Explain the difference between data replication in a distributed system to the maintenance of a remote backup site? Compare the advantages and disadvantages of the data replication and data fragmentation approaches in a distributed system. When is it useful to have replication or fragmentation of data?

2. What is the major reason to cause false deadlock in the distributed databases? Please provide an example of false deadlock which is different from the one shown in the lecture slides.

3. Consider the relations:

*employee*(name, address, salary, plant\_number)  
*machine*(machine\_number, type, plant\_number)

Assume that the *employee* relation is fragmented horizontally by *plant\_number*, and that each fragment is stored locally at its corresponding plant site. Assume that the *machine* relation is stored in its entirety at the Armonk site. Describe a good strategy for processing each of the following queries.

- a. Find all employees at the plant that contains machine number 1130.
- b. Find all employees at plants that contain machines whose type is “milling machine.”
- c. Find all machines at the Almaden plant,
- d. Find employee  $\bowtie$  machine.

4. For each of the strategies of problem 3, state how your choice of a strategy depends on:

- a. The site at which the query was entered.
- b. The site at which the result is desired.

5. In a mediator (data integration) system, the mediated schema is

Customer (ID, name, statecode)  
State(statecode, statename)

There are two data resources:

S1:  
Customer (ID, full\_name, region)

S2:

Customer (ID, first\_name, last\_name, region\_code)

Region (region\_code, region\_name)

Assume “name” in the mediated schema means “full\_name” in S1. “Region” in data resources means “state” in the mediated schema. For example, “state-code” or “region\_code” of “Oregon” is “OR.” Can you do schema mediation (mapping) in Global-as-View or Local-as-View, or both? Write down your solution(s) in SQL views.

Write a query “List all customer names in ‘Oregon’” in SQL based on the mediated schema. Can the query be answered with your views?

---

**To turn in by paper version:** Ask Cheri to put your answers to Prof. Dejing Dou’s mailbox.

**To turn in by email:** Please email your answers to [dou@cs.uoregon.edu](mailto:dou@cs.uoregon.edu). We prefer that you send in a pdf file. If you are using Word, you should convert your word file to a pdf file.