Assignment 3
CIS 670 Data Science, Winter 2019
due 11:59 pm, Monday, February 11th

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

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

There are two data resources:

S1:

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

S2:

$Customer(ID, full\_name, region)$

Assume "name" in the mediated schema means "full_name" in S2. "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?

2. Prove the following statement: Given two LAV data integration systems $\mathcal{I}_1 = \langle \mathcal{G}, \mathcal{S}_1, \mathcal{M}_1 \rangle$ and $\mathcal{I}_2 = \langle \mathcal{G}, \mathcal{S}_2, \mathcal{M}_2 \rangle$, $\mathcal{I}_1$ is **p-contained** in $\mathcal{I}_2$ if, for each query Q, $cert_{[Q,\mathcal{I}_1]}$ equivalent to Q implies $cert_{[Q,\mathcal{I}_2]}$ equivalent to Q.

3. Give a real world multimodal data fusion system which is not mentioned either in the lecture or in the survey paper by (Atrey *et al.* 2010).

4. What is the relationship between Deep Learning and Big Data? Give a real world Big Data application which Deep Learning technique(s) show advantages than other traditional machine learning techniques (e.g., Bayes Networks, SVM), why? Give an example Deep Learning technique or application that Big Data help optimize or improve the performance, why?

To turn in by emails: Email your answers to **dou@cs.uoregon.edu**. A pdf file is preferred. If you are using Word, you should be able to convert your word file to a pdf file.

To turn in by paper version: Ask Jon or Cheri to put your answers to Prof. Dejing Dou's mailbox.