The assignment is to be turned in before Midnight (by 11:59pm) on March 15th, 2018. You should turn in the solutions to this assignment as a pdf file through the TEACH website. The solutions should be produced using editing software programs, such as LaTeX or Word, otherwise they will not be graded.

1: Two Phase Commit (2 points)

Explain the actions taken in the two phase commit protocol after restart if the coordinator or subordinate nodes fail.

2: Multi-version concurrency control (2 points)

Consider transactions T_1, T_2, T_3 with timestamps 10, 20, and 30, respectively in a database system that follows multi-version concurrency control protocol. T_1 is a reader and T_2 and T_3 are writer transactions.

(a) Let T_1 read and T_2 read and update data item D. If transaction T_3 updates only D, provide a schedule where the multi-version concurrency control method will **not** restart T_3.

(b) Again, assume that T_1 reads and T_2 and T_3 read and update only data item D. Provide a schedule where the multi-version concurrency control method will restart T_2.

3: Quorum reading (2 points)

Let the number of copies of a data item in a cluster, N, be 10 and the number of nodes that participate in a successful write, W, be 5.

(a) What is minimum number of nodes that should participate in a successful read, R?

(b) Now, consider a sloppy quorum with the same values for N and W. What is the minimum number of nodes that should participate in a successful read, R?

4: Eventual consistency and vector clocks (2 points)

Each row in the following table shows vector clocks of different copies of the same data on a system with three nodes S_X, S_Y, and S_Z. Explain if the copies in each row have a conflict.

Copy 1	Copy 2
($[S_X,10],[S_Y,20],[S_Z,10]$)	($[S_X,10],[S_Y,20]$)
($[S_X,10],[S_Y,10],[S_Z,10]$)	($[S_X,10],[S_Y,20]$)

Table 1: Vector clocks

5: PageRank (2 points)

Assume the following graph depict a part of the Web, where nodes represent pages and edges show hyper-links. Find out the pages whose PageRank values are greater than zero and their relative PageRank values in the graph. You do **not** need to perform the fix point computation to determine the PageRank values. Instead, you should guess the PageRank values based on your understanding of the PageRank algorithm and explain why you think they are correct. If it is no possible to make any educated guess for some page(s), you should explain why.
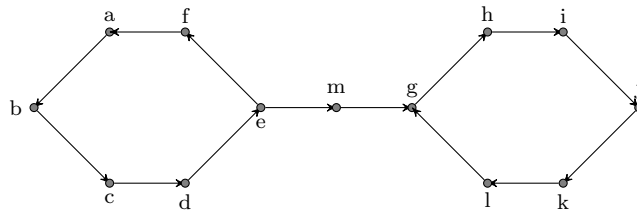

Figure 1: The graph for Problem 5

6: MangoDB (6 points)

In this part, you will work with a MongoDB database. In order to have access to your database, you should be on campus, use OSU VPN or SSH to an engineering server. First you need to download and import the Restaurants database to your DBMS. You can download the database from *https://raw.githubusercontent.com/mongodb/docs-assets/primer-dataset/primer-dataset.json*. Then, you should use the following command to import the data (username, password and db-name is `cs440_<your_onid>`):

```
mongoimport --db <db-name> --collection restaurants --drop
  --file <address to primer-dataset.json> --host classmongo.engr.oregonstate.edu
  -u <username> -p <password>
```

Next connect to your MongoDB using the following command:

```
mongo -u <username> -p <password> classmongo.engr.oregonstate.edu/<db-name>
```

It is recommended that you change your password the first time you connect to the database. You can use the following command to change it:

```
db.changeUserPassword(cs440_<your_onid>,newpass)
```

Next you should go over the tutorial of the commands that you will need in this assignment. For part a, you need to use the *find* command. You may find a tutorial for it at *https://docs.mongodb.com/getting-started/shell/query*. For part b & c you need to do text search explained at *https://docs.mongodb.com/v3.2/text-search* and for the last part go through this tutorial *https://docs.mongodb.com/getting-started/shell/aggregation*. After connecting to the database execute the *findOne()* command to see a sample document. This will help you to understand the structure of the documents.

1. Write a query that returns the cuisine type of the restaurant named The Dead Rabbit (1 point).

2. Write a query to create an index on the name attribute of the restaurants (1 point).

3. Write a query that uses the index created in part b to return all the restaurants containing the term Rabbit in their name (2 points).

4. Write an aggregate query to show total count of restaurants in each borough (2 points).