CS 488/588 Cloud & Cluster Data Management  
Fall 2019 Quarter

Project Part 2: Data Modeling and Application Design - DRAFT

*Due:* Thursday, Nov 14, by midnight (D2L)

***Part 2 Goal:*** Design the model/storage of the data and the implementation of the specified queries given for the system the team investigated in Part 1 of the project.

***Deliverable:*** The deliverable for this part will be a short writeup of your design and likely be a meeting with the Prof & TA to discuss your design. Exact deliverable to be determined. Writeups will be turned in on D2L

***Data Sets:*** There are two options for the data and related questions to be used for the project. DBLP and Freeway. Both data sets are in the Class Google Drive Folder. Both data sets contain data and a set of questions to be answered over the data.

*Options 1: DBLB (Graph Data)*

This option focuses around an application that manages and processes a graph of bibliographic data derived from the DBLP Computer Science bibliography: <http://dblp.uni-trier.de/>

*Option 2: Freeway Data (Relational Data)*

This option focuses around freeway speed data and calculating travel times. The data set is relational.

**Data Modeling & Application Design**

In the second part of the project, teams design the storage/modeling of the provided data and designs how they will implement the questions provided for the selected data set. All for the system the team profiled in Task 1.

The deliverable for this part of the project is a brief report (submitted on D2L) that includes the items below. In addition, there will likely be a meeting with the Prof and TA to discuss your design.

* Using the provided example data, explain how you store the data structures in your data store.
  + Show how the example data can be represented in the data model of your data store. Mention any indexes you will create.
  + Present at least one variant you considered during the design phase.
  + Explain why you believe your final design is the preferred one
* Describe the process you will use to restructure the data to fit your model.
* Outline implementation strategies in pseudo-code based on the capabilities of your data store and the indexes you plan to define for all of the provided questions.
* Include proof that you have loaded at least a few data items into your system.
* Discuss how your design could handle updates to the data.

Note: For Task 2, you will design for all 6 questions. For Task 3, you will select 4 of the 6 questions to implement.