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

Project Part 3: Prototype Implementation

*Due:* Presentations during the last week of class and final exam slot. Individual write-ups due on D2L by midnight Dec 10.

Part 3 will be done in teams; however, the 2-page write-up will be done individually.

**Task 3: Prototype Implementation**

Briefly, Task 3 is to implement your design from Task 2. More specifically, based on the design from Task 2, your group should create a prototype in your chosen system that implements your data model, loads the DBLP data and implements (some of) the queries from Task 1. Each student will also do an individual short project report.

**Deliverables:**

1. Option 1: A 10-15 minute presentation on your prototype. Please upload your presentation slides to the Class Google Drive folder. The presentation should include:
   1. Demonstration of at least 2 queries. (The slides should include documentation that you have at least 4 queries running, please discuss 2 in the presentation.)
   2. Critique of your data model and query execution plans. What would you change if you could do the design over? What did you change?
   3. Lessons learned: What did you learn about your system (and cloud data management in general) from the project? What advice would you give someone who was considering using the system you used?
   4. Graphics are strongly encouraged. Graphics that effectively communicate your project and results will increase your grade.
2. Option 2: A ~10 page paper, minimum 11 pt font, 1.2 space. (If you are a group of 1 or 2, you can request a shorter length paper.) The paper should include:
   1. Documentation that you have at least 4 queries running.
   2. Critique of your data model and query execution plans. What would you change if you could do the design over? What did you change?
   3. Lessons learned: What did you learn about your system (and cloud data management in general) from the project? What advice would you give someone who was considering using the system you used?
   4. Graphics are strongly encouraged, you could target 5 pages of writing and 5 pages of graphics. Graphics that effectively communicate your project and results will increase your grade.
3. A copy of your code. Ideally the code is submitted as a github link.
4. A 2-page report, done individually, that summarizes the major decisions made, gives lessons learned, and explains your role in the project.

**Minimum Requirements and Scoring Details:**

* Your individual report is 40% of the score.
* You should implement at least 4 out of the 6 queries. You can get up to 10% extra credit for doing more.
* You can work with a subset of the data. If you do so, explain what subset in your slides. There can be up to a 10% reduction if you use less than the full data set. If you are using your own data set, please aim for ~700MB of (raw) data.