

# CS 410/510 Data Management in the Cloud

## Spring 2018 Quarter

---

### Project Part 4: Prototype Implementation

Due: Presentations during the final exam slot on Tuesday, 12 June 2018, 10:15a.

Code must be submitted by 5p Tuesday, 12 June.

Individual write-ups due at presentation time.

You should do this assignment in project groups (but the 2-page write-up should be done individually). You should only talk to your group, the TA and the instructor about this assignment. You may also post questions and comments to the course discussion list on Piazza.

### Task 4: Prototype Implementation

Briefly, Task 4 is to implement your design from Task 3. More specifically, based on the design from Task 3, 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. A 10-minute presentation on your prototype. ***Please bring a hardcopy of the slides to the presentation.*** The presentation should include:
  - a. Demonstration of at least 2 queries.
  - b. Critique of your data model and query execution plans. What would you change if you could do the design over? What did you change?
  - c. 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?
2. A copy of your code. It should be commented and submitted as a PDF via Piazza. Please include all group members' names in the submission.
3. 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.