

# CS 487/587 Database Implementation

## Spring 2019

### Database Benchmarking Project

---

#### Project Goals:

- Increased understanding of Database Performance
- Get experience with a relational-style data management system of your choice
- Programming Experience

#### Project Part I: Data Generation & System Selection

**Due:** Monday April 22, midnight, D2L

**Teams:** This project is intended to be done in teams of two. Working individually is acceptable.

**Turn-ins:** To turn in your results, please create a GitHub repository that includes the following. Then turn in that link on D2L.

- Your data generation code
- Sample data files to demonstrate that you have successfully generated data (~50 rows for each table)
- README-Part1 file. 1-2 pages including:
  - Brief description of your work for Part 1
  - Which system you will be working with and why you chose it
  - Demonstrate you have loaded data into that system
  - Include lessons learned or issues encountered

#### Details:

**Data Generation:** Write scripts to generate data based on the Wisconsin Benchmark specification as described in the paper posted on D2L. You may use the programming language of your choice. Data may be generated in csv format; or you may generate and load data directly into a system if you prefer. The data generation code must be your own work. If you look at other code, you must reference it.

**System Selection:** Select a system to use for your project. The core requirement is that the system you select must have a SQL interface and you are required to use that SQL interface (and only the SQL interface) for this project. Suggested systems: postgresql (hosted on campus), BigQuery (GCP)

**Load Data:** Load a small amount of data into the system.