CMPE 266– Big Data Engineering and Analytics

Spring 2023

Group Project

Group project guidelines and deadlines

Projects ideas should be discussed with the instructor outside the class to negotiate the scope and define the project requirements and grading rubric. Sign-up for half hour slots Apr 13-14. Use this link to sign up: https://calendar.google.com/calendar/selfsched?sstoken=UUNMYmFrb2pnajVQfGRlZmF1bHR8OThiOGQ1ZDZjMmUyMGNiZDAyOWYwOGViNzJhZGM2YzM

Some general ideas for projects:
• Implement a system using a NoSQL datastore as the backend (not relational) or flat files (for non-structured data). Interface should be preferably web-based. You are not required to implement a whole system but the parts that showcase the main capabilities/functionality of the system. For example, provide a form for entering data and performing searches. Some sample apps: implement an address book, a points-of-interests application (using maps for the interfaces), a search-engine to query documents by example and retrieve similar documents to the query.

• Index a large data collection using a multi-dimensional index described in class and perform NN queries or data mining analysis (clustering, classification, outlier detection). Projects of this type should include analysis of the index used (varying parameters, comparing performance with no-index approach, measuring cost of the index).

Total Project: 20 points
• **Project Proposal** (2 points) – One page overview describing the project you'll develop. Due Apr,18. On Canvas.

• **Project Progress Report** (3 points) – A detail description of the project, project milestones, and tasks completed. Due May 2nd, by 11pm. On Canvas.

• **Project Presentation** (8 points) – In class presentations: results and/or demo (10-minute presentations) – In class May 11 (Schedule TBD).

• **Project Report** (7 points) – Includes description of the tasks completed and documentation. Your code should be provided as a github repository. The link to the code repository shall be included in the report. It should also include complete code and instructions to execute the project, as well as test data (if needed).