

|  |  |
| --- | --- |
| **Course:** | CS6020 |
| **Assignment:** | Term Project |
| **Total Points:** | 100 |
| **Date Due:** | Posted on Blackboard |

# Learning Objectives

In this assignment, you will learn how to:

* combine the technologies, concepts, and strategies learned in this course
* apply your new knowledge

# Project Goal

The term project is intended to allow you to revisit what you have studied this semester and deepen some aspect of the material based on your interests. The project may be a programming project, data storage design, technology review, research paper, or some other "capstone" project.

Project Proposal

You must submit a proposal of what you plan to do. The proposal must be posted on the Blackboard Discussion Forum under the "Term Project" discussion thread. You must also reply to EVERY other submitted proposal providing your feedback. The feedback must be taken into account when developing the final submission. The Blackboard post should be 1 - 3 paragraphs clearly outlining what you plan to do, how to plan to approach it, and what your final deliverable will be. As you are making progress, update your thread on Blackboard and keep contributing to the other threads.

Project Ideas

Here are some ideas for a term project; you are not limited to those ideas and may propose anything else that would be about 30-60 hours of work.

* Research paper on a database technology, data quality, data cleaning
* Review of the literature and summary of at least five related research papers published in the past three years in peer-reviewed venues (journals, conferences, symposia)
* Implementation of a database using a data set from your work
* Loading and cleaning of a data set from work with storage in a database
* Review of NoSQL databases and when/how they would be used or are most applicable
* Implementation of a prototype data store in a NoSQL database with R, such as Hadoop, Cassandra, CouchDB, MongoDB, Redis, Riak, etc.

Not every idea will get maximum credit. A maximum credit project is one that takes a data set from some source, cleans it, stores it in a NoSQL database, and retrieves the data and shows some analysis of the data.

Deliverable

You must submit a detailed and well-formatted report in PDF format that is free of grammatical errors, contains proper references, and shows the work you did. You must clearly state the problem and show how you solved it. You should also include any issues you ran into and how you resolved it. You must show outputs from your R code and screenshots of it running. The report must be detailed and should allow one to judge what you did and how well you did it.

Your report must conclude with insights on what you learned.

A good report should be about 10 pages and show clearly what you did. Your report should have a title page with your name, course number, and the semester. The first page of the report must be your original proposal (directly copied from Blackboard.)