1. Introduction

This document describes the basic logistics required to successfully complete the term project and term paper for MSDS 7330. In performing the term project, the students will draw upon their knowledge and experience from the course lectures, readings and professional experience.

2. Scope , Logistics and Timelines

2.1. **Logistics and Timelines.** The work entails one project proposal , one final report and an oral presentation due in conjunction with the project.

The project proposal, due on 08/08/2018 is a one/two page write up describing your project , the tools and the technologies you intend to use and the team you will be working with. The proposal will contain

(1) Project title
(2) Names and email addresses of the team members
(3) Clear statement of the problem : A one to two sentence statement of the problem followed by a one paragraph clarification of the problem. The paragraph should identify clearly the question you are addressing, and proper motivation for the importance of the problem must be provided.
(4) Clear statement of your design/development methodology. A brief explanation of how you will approach and develop a solution to the problem
(5) A list of tools , technologies and resources needed to accomplish your work,with special emphasis on important pieces you dont yet have access to and your plan to overcome the same.

Note: Email your proposals to rsrinivas@smu.edu with the following subject line:
MSDS7330 Summer2018 : Term project proposal
One email per group ONLY!

The final report,due on 08/20/2018 is a power point presentation describing the design and implementation of your project The oral presentation, is a final presentation given during the last week of classes(08/20/2018). All students are encouraged to attend the last week of classes. Reach out in advance to your instructor for any exceptions.

2.2. **Scope.** Good class projects can vary dramatically in size, complexity, scope and topic. The only real limitation is that the project must relate in some way to the course. Note that purely simple problems, such as installing a large distributed database system, while interesting and potentially a source of great learning, are not acceptable projects. You can identify a project related to your work, so that you may have a ready platform on which to apply your knowledge and put your skills to work immediately Alternately you can choose from the list of sample projects listed at the end of this document.

2.3. **Teams.** Teams of two or three are encouraged for the term project (teams of one are also acceptable). Teams typically accomplish more and learn more than do persons working alone, so start looking for teammates as soon as possible!

3. EXAMPLE PROJECT TOPICS

Note: These are only references. Your project may be different from the ones listed below.

(1) Asynchronous Database Access. Client software interacts with standard SQL databases via a blocking interface like ODBC or JDBC; the client sends SQL, waits for the database to process the query, and receives an answer. A non - blocking interface would allow a single client thread to issue many parallel queries from the same thread, with potential for some impressive performance gains. This project would investigate how this would work (do the queries have to be in different transactions? what kind of modification would need to be made to the database) and would look at the possible performance gains in some typical database benchmarks or applications.

(2) Data at rest refers to any data that is stored physically in any digital form, like the data stored within a database. Many applications transact with sensitive user information such as SSNs , credit card numbers etc. Encrypting such data is vital for consumer safety and security reasons. This project will study and present the various encryption options available in MySQL and MongoDB. The project will also entail simulating data encryption on sample data in either MySQL or MongoDB

(3) Client side database. Build a Javascript library that client side Web applications can use to access a database; the idea is to avoid the painful way in which current client side application have to use the XMLHttpRequest interface to access server side objects asynchronously. This layer can be enabled to cache objects on the client side whenever possible, but be backed by a shared server side database system.

(4) Twitter provides a fire hose of data. Automatically filtering, aggregating, analyzing such data can allow a way to harness the full value of the data , extracting valuable information. The idea of this project is investigating stream processing technology to operate on social streams. The project will also entail building a sample service that accepts data from a twitter stream and stores it locally in a database. The project report will discuss the different design considerations such as real-time v/s batch processing , performance , scalability etc.

(5) Teach the class about a topic not covered by one of the units, such as Cassandra, Redis or any of the numerous other topics we have not covered. The presentation will incorporate lecture slides and a demonstration of the underlying software / tools being taught. Prepare a video presentation that may be provided asynchronously to the class. Note: Only 2 per team allowed for such projects.