Dear CS257 Classes:

In this big data era, do we still need to study traditional databases? Please look at the article Communications of the ACM, January 2015, pp 18. (see http://xanadu.cs.sjsu.edu/~drtylin/classes/cs157A_fall2015/notes/ by Michael Stonebraker (The Turing Award Winner of 2014) : " ¼ a new DBMS, IMPLA, which runs on HDFS (Hadoop File System). Put simply, IMPLA is architected exactly like all of the share-nothing parallel SQL DBMS, serving the data warehouse (see Ch 21.2.2) market. Specifically, notice the MapReduce Layer has been removed, and for good reason" …. "Hadoop is picking up support in the general community." "a delay for a decade." "Goole has long since abandon it."

1) We will cover DBMS in lectures,
2) The so-called "No-SQL" will be covered in the Project.

This green sheet is prepared for "standard course"; this semester; we will increase the weight of the project.

1) Please down load the project based on the instructions giving
   http://xanadu.cs.sjsu.edu/~drtylin/classes/cs157A_fall2015/projects/
2) Please read the file Google_Big_Data_Decade.pdf in
   http://xanadu.cs.sjsu.edu/~drtylin/classes/cs157A_fall2015/notes/

Dr. Lin
http://dl.acm.org/inst_page.cfm?id=60015609

About the Project:

1) You have to sign a non-disclosure form
2) The project is the core of a semantic based search engine
   Which is a project in Data Science
   Vasant Dhar**, PhD,** Editor**-in-**Chief **of** Big Data (A professor of NYU) stated:
   "Data Science is a study of generalizable extraction of knowledge from data."
3) It is the product of CS298-99, since 2005 (Albert Sutojo he computed TFIDF using database concept), Important observation by Tam Ngo in 2006: Google's latent sematic index has nothing to do with semantic. The best version is by Jean David Hsu (an undergraduate); he tested on 3 million abstracts from Medline. Current version is by Bieu Do. A secondary storage version by Richard Deeley (undergraduate)
4) Your will be assigned

San José State University
Science/Computer Science
CS 257 Database Management Systems I

1. Course Information

Instructor: Dr. Tsau Young Lin
Department of Computer Science
College of Science, San Jose State University.
Spring Semester, 2015

Course Title	**Principles of Database System – Projects in Big Data**					
Course No	CS257					
	Section No	start	end	days	building	rooms
	1	10:30	11:45	MW	MH	225
	2	10:30	11:45	TR	MH	222
Office Hours:	MW14:45-15:30; 11:45-12:15 (appointments only)					
Office	MH214					
Office Phone:	(408)924-5121					
E-mail:	ty.lin@sjsu.edu					
Department Fax:	(408)924-5062					

2. Specific course information:

CS257 is different from CS157A/B. In CS157A/B you learn how to use database and extended systems, but in CS257 you will learn how database and extended systems work internally.

Catalog Description:

> (1) Design management and performance issues on: file organization and access methods, buffer management and storage management. (2) Query processing and query optimization, transaction management, recovery, and (3) concurrency control techniques. Reliability, (4) protection and integrity techniques. (5) Extensive programming project. Prerequisite: CS 157B or instructor consent.

University Rules
NOTE that University policy F69-24 at http://www.sjsu.edu/senate/docs/F69-24.pdf states that "Students should attend all meetings of their classes, not only because they are responsible for material discussed therein, but because active participation is frequently essential to insure maximum benefit for all members of the class. Attendance per se shall not be used as a criterion for grading." Note that "All students have the right, within a reasonable time, to know their academic scores, to review their grade-dependent work, and to be provided with explanations for the determination of their course grades." See University Policy F13-1 at http://www.sjsu.edu/senate/docs/F13-1.pdf for more details.

Specific goals for my sections:

The focus is on project; each team has different schedules; there are no class schedules

> The impact of Big Data to CS257 is tremendous, so this semester; we focus on project that continues (1) the building of a text/web based non SQL DBMS that was built in CS257-267 for several years (Big Data DBMS). In this part (1) and (2) will be addressed; note that the results of text/web analysis are stored in MS SQL. (3) Concurrent control are replaced by

Map-Reduce; (4) due to time limit, we do not plan to cover protection. However, we will do extensive data analysis that may address some integrity issues. In addition, a new data analysis technique that are derived from algebraic topology will be explained.

a. Programming outcomes:
   1) to use commercially available DBMS, such as Oracle, DB2, MS SQL Server and etc (we will use MS SQL Server)
   2) to handle the interactions between DBMS and a third generation programming language, such as Java, C++.
   3) Ability to organize a team project to build a "real world" modern database system(e.g., non-SQL DBMS)

3. Textbook

Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer D. Widom.
Database Systems: The Complete Book. Prentice Hall. 2nd Ed ISBN-13: 978-0131873254. 2008.

4. Supplemental reference:

Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems) By Jiawei Han and Micheline Kamber, 2012

Data Mining : Introductory and Advanced Topics
by Margaret D. Dunham
Publisher : Prentice Hall, 2002
ISBN : 0130888923

Research papers on Rough Sets and Granular Computing
(Rough set theory is a name for algebraic theory of relational databases)

5. Student learning objectives for the course:

Upon successful completion of this course, students should have overall knowledge of logical views of data, physical storage structuring techniques, related access methods and query optimizations for classical database systems, and have a good idea about the important issues when the systems are extended to bio, social and big data.

6. Tentative course calendar including assignment due dates, exam dates, date of Final exam

Course Plan

Weeks	Lectures & Assignments
0(8/17)	Overview of the class, policies(Green sheet) Project Description
1(8/24)	Query Compiler (Introduction and Query Algebra)
2(8/31)	Query Compiler (Introduction and Query Algebra)

3(9/7)	Query Compiler (Introduction and Query Algebra)
4(9/14)	Secondary Storage Management * Power pointer file due (2/28)
5(9/21)	Secondary Storage Management
6(9/28)	Secondary Storage Management
7(10/5)	Secondary Storage Management
8(10/12).	Concurrency Control
9(10/19)	Concurrency Control
10(10/26)	Concurrency Control
11(10/31)	Query Execution
12(11/2)	Query Execution
13 (11/9)	Query Execution
14(11/16)	Project demo
15(11/23)	Project demo
16(11/30)	Project demo
17(12/7)	Review and Discussions
18(12/10-16)	Final Exam

8. Grades

Projects & home works	30%
Exams	30%
Quizzes(Class average set 80 linearly)	10%
Final Exam	30%
Total	100%
90-92; 93-96;97-	A

80-82; 83-86;87-	B
70-72; 73-76;77-	C
60-62; 63-66;67-	D
<60	F

NOTE that University policy F69-24 at http://www.sjsu.edu/senate/docs/F69-24.pdf states that "Students should attend all meetings of their classes, not only because they are responsible for material discussed therein, but because active participation is frequently essential to insure maximum benefit for all members of the class. Attendance per se shall not be used as a criterion for grading." Note that "All students have the right, within a reasonable time, to know their academic scores, to review their grade-dependent work, and to be provided with explanations for the determination of their course grades." See University Policy F13-1 at http://www.sjsu.edu/senate/docs/F13-1.pdf for more details.

9. University Policies

General Expectations, Rights and Responsibilities of the Student

As members of the academic community, students accept both the rights and responsibilities incumbent upon all members of the institution. Students are encouraged to familiarize themselves with SJSU's policies and practices pertaining to the procedures to follow if and when questions or concerns about a class arises. See University Policy S90–5 at http://www.sjsu.edu/senate /docs/S90-5.pdf. More detailed information on a variety of related topics is available in the SJSU catalog, at http://info.sjsu.edu/web-dbgen/narr/catalog/rec-12234.12506.html. In general, it is recommended that students begin by seeking clarification or discussing concerns with their instructor. If such conversation is not possible, or if it does not serve to address the issue, it is recommended that the student contact the Department Chair as a next step.

Dropping and Adding

Students are responsible for understanding the policies and procedures about add/drop, grade forgiveness, etc. Refer to the current semester's Catalog Policies section at http://info.sjsu.edu/static /catalog/policies.html. Add/drop deadlines can be found on the current academic year calendars document on the Academic Calendars webpage at http://www.sjsu.edu/provost/services /academic_calendars/. The Late Drop Policy is available at http://www.sjsu.edu/aars/policies/latedrops /policy/**.** Students should be aware of the current deadlines and penalties for dropping classes.

Information about the latest changes and news is available at the Advising Hub at http://www.sjsu.edu /advising/.

Consent for Recording of Class and Public Sharing of Instructor Material

University Policy S12-7, http://www.sjsu.edu/senate/docs/S12-7.pdf, requires students to obtain instructor's permission to record the course and the following items to be included in the syllabus:

- "Common courtesy and professional behavior dictate that you notify someone when you are recording him/her. You must obtain the instructor's permission to make audio or video recordings in this class. Such permission allows the recordings to be used for your private, study purposes only. The recordings are the intellectual property of the instructor; you have not been given any rights to reproduce or distribute the material."
    - It is suggested that the greensheet include the instructor's process for granting permission, whether in writing or orally and whether for the whole semester or on a class by class basis.
    - In classes where active participation of students or guests may be on the recording, permission of those students or guests should be obtained as well.
- "Course material developed by the instructor is the intellectual property of the instructor and cannot be shared publicly without his/her approval. You may not publicly share or upload instructor generated material for this course such as exam questions, lecture notes, or homework solutions without instructor consent."

Academic integrity

Your commitment, as a student, to learning is evidenced by your enrollment at San Jose State University. The University Academic Integrity Policy S07-2 at http://www.sjsu.edu/senate /docs/S07-2.pdf requires you to be honest in all your academic course work. Faculty members are required to report all infractions to the office of Student Conduct and Ethical Development. The Student Conduct and Ethical Development website is available at http://www.sjsu.edu /studentconduct/.

Campus Policy in Compliance with the American Disabilities Act

If you need course adaptations or accommodations because of a disability, or if you need to make special arrangements in case the building must be evacuated, please make an appointment with me as soon as possible, or see me during office hours. Presidential Directive 97-03 at http://www.sjsu.edu /president/docs/directives/PD_1997-03.pdf requires that students with disabilities requesting accommodations must register with the Accessible Education Center (AEC) at http://www.sjsu.edu/aec to establish a record of their disability.