

San Jose State University
College of Science
Department of Computer Science
CS185C, Introduction to NoSQL databases, Spring 2017

Course and Contact Information

- Instructor: Dr. Kim
- Office Location: MacQuarrie Hall 217 (MH217)
- Telephone: 408-924-5122
- E-mail: suneuy.kim@sjsu.edu (Preferred mode of contact is via email.)
 - When you send me an e-mail to ask a question, use [Q] in a subject line to get a reply from me within a reasonable response time. Here is an example subject line to ask a question.

[Q] lecture note

- Office Hours: TW 9:15 - 10:15 MH 217
- Class Days/Time/Classroom
 - Section 2 (Lecture): TR 10:30 - 11:45 MH 233
- Course Prerequisites: CS157A equivalent
- [Course Web Site](http://www.cs.sjsu.edu/~kim/nosql) at <http://www.cs.sjsu.edu/~kim/nosql>
Announcements and course materials will appear here. It is updated frequently. You are strongly encouraged to check out this course web page regularly.

Course Description

NoSQL databases emerged to meet a new set of technology requirements imposed by modern enterprise applications. The course covers the features and data models of NoSQL databases, representative NoSQL databases and their use cases, data processing in NoSQL databases using Hadoop and Apache Hive, and programming using API for NoSQL databases. A team project is required which involves configuration and deployment of a NoSQL database of choice, populating the database with publicly available datasets, and API coding.

Course Learning Outcomes

Upon successful completion of this course, students should be able to:

- configure and deploy NoSQL databases on their laptop
- populate NoSQL databases with publicly available datasets
- develop a moderate-size database application using API coding with a back-end NoSQL database
- choose NoSQL databases based on use cases for future development work

Course Topics

Topics	Weeks
Why NoSQL?	0.5

Features of NoSQL Databases	0.5
NoSQL Data Models	1.0
NoSQL Distribution Models	0.5
MongoDB	3
API coding with MongoDB	0.5
Cassandra	3
API coding with Cassandra	0.5
NoSQL in the cloud: Amazon AWS	0.5
Processing NoSQL data: Map-Reduce, Hadoop Hive	0.5
Project proposal presentations	1.5
Paper presentations	2
Total	14

Required Texts/Readings

- Textbook: None required
- References (most of the following references are available at the Kings library in a form of e-book)
 - NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence by Parmod J. Sadalage and Martin Fowler (not available in the Kings)
 - MongoDB: The Definitive Guide: Powerful and Scalable Data Storage 2nd Edition by Kristina Chodorow, May 2013
 - The Definitive Guide to MongoDB: A Complete Guide to Dealing with Big Data using MongoDB, Third Edition by David Hows, Peter Membrey, Eelco Plugge and Tim Hawkins, December, 2015
 - Mastering Apache Cassandra - Second Edition by Nishant Neeraj, March 26, 2015
 - Cassandra: The Definitive Guide: Distributed Data at Web Scale Jul 22, 2016 by Jeff Carpenter and Eben Hewitt
- Other readings: A list of references will be provided as topics covered on the course web site

Course Requirements and Assignments

- Programming Assignments, Project, and Presentation
 - Programming Assignments: 4~5 individual programming assignments are given, unless otherwise specified.
 - Team Project
 - A team of three people conducts the project.
 - The project involves configuring and deploying a NoSQL database, data population, and programming using API.
 - There will be a 15 minute-long project proposal presentation per team.
 - The final result of the project will be submitted through the project submission link on the course web site.

- Team Presentation
 - A team of three people will study an assigned paper and give a 30 minute-long presentation.
 - The team will choose a paper from the list which will be prepared by the instructor.
- Submission/Late Policy
 - Any assignments/project turned in past the deadline will get a penalty: For each late day, a 20% of the maximum obtainable score of the work will be taken out of what you earned. (a late day is one 24 hour period beyond the due date). For example, suppose the maximum score of an assignment is 100 and you earned 80 points. If the submission is late by two days, the final score of the assignment would be $80 - 2 * 20 = 40$.
 - Any submission turned in more than 48 hours past the deadline will result in a grade of zero for that assignment.
 - On-line submission: You can submit your work multiple times. If then, the latest one will be considered as the final submission. If the final submission is late, the late policy will be applied.
 - E-mail submissions will not be accepted for grading.
- Teamwork Policy
 - Once a team is formed, it will last throughout the semester. If you dissolve your team, a significant amount of penalty will be determined by the instructor and given to both parties.
 - For the project, students are expected to submit their peer evaluation in addition to the final report. The task responsibility and contribution of every team member must be precisely documented in a peer evaluation form.
- Software
 - MongoDB
 - Cassandra
 - Oracle Virtual Box
 - Linux (Ubuntu, CentOS, etc)
 - Programming Language: Java and/or Python

Evaluation (Exams)

- There will be one midterm exam and one comprehensive final exam. The exams are scheduled as below. The dates of midterm exams are subject to change with fair notice, but the final exam date is firm and cannot be changed.
 - Midterm Exam: Thursday, March 23 in class
 - Final Exam: Monday, May 22 0945-1200
- Makeup Exam Policy Absolutely no make-up exams will be offered under any circumstances. For those who couldn't take the exam or worked hard but had a bad day on the exam day ending up with a low score, I offer the following opportunity to possibly replace your worst midterm score with the final score. If your final exam (percentage) grade is higher than your worst midterm (percentage) grade, then I will replace the worst midterm grade with your final exam grade. For example, if you have a 60% on your worst midterm and you receive an 80% on the final exam, I will replace the 60% by 80% in the computation of your course grade.

Grading Information

You will receive the final grade based on the weighted average score on your performance. The grading weights are as follows.

- Assignments: 20%
- Midterm: 20%
- Final Exam: 25%
- Project: 20%
- Presentation: 13%
- Participation: 2% (participating in project presentation sessions)
- First I try scores of 90, 80, and 70 to cutoff letter grades of A-, B-, and C-, respectively. If overall class performance is too low to use these cut offs, I set a cut off of C- to a lower score than the class total average but a higher score than 60 (this number may change), and divide the students' group above the cut off of C- into A+, A, A-, B+, B, B-, C+, C, C-. The rest of students will be given by a grade of D+, D, D-, F or WU depending on their class performance.
- The same method will be applied to every student enrolled in the class including graduate students.

Classroom Protocol

Consent for Recording of Class and Public Sharing of Instructor Material

: [University Policy S12-7](http://www.sjsu.edu/senate/docs/S12-7.pdf), <http://www.sjsu.edu/senate/docs/S12-7.pdf>, requires students to obtain instructor's permission to record the course:

- "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."
- "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."

University Policies

Per University Policy S16-9, university-wide policy information relevant to all courses, such as academic integrity, accommodations, etc. will be available on [Office of Graduate and Undergraduate Programs' Syllabus Information web page](http://www.sjsu.edu/gup/syllabusinfo/) at <http://www.sjsu.edu/gup/syllabusinfo/>

CS185C: Introduction to NoSQL, Spring 2017: Semester Schedule

Subject to change with fair notice.

Week	Dates	Topics	Assignments
1	1/26	Course Orientation	
1	1/31	Fundamentals of NoSQL	

2	2/2	Fundamentals of NoSQL	
2	2/7	Fundamentals of NoSQL	
3	2/9	Fundamentals of NoSQL	
3	2/14	MongoDB	
4	2/16	MongoDB	
4	2/21	MongoDB	
5	2/23	MongoDB	
5	2/28	MongoDB	
6	3/2	MongoDB	
6	3/7	API Coding for MongoDB	
7	3/9	Project Proposal Presentation	
7	3/14	Project Proposal Presentation	
8	3/16	Project Proposal Presentation	
8	3/21	Cassandra	
9	3/23	Midterm	
	3/28	SPRING RECESS	
	3/30	SPRING RECESS	
9	4/4	Cassandra	
10	4/6	Cassandra	
10	4/11	Cassandra	
11	4/13	Cassandra	
11	4/18	Cassandra	
12	4/20	Cassandra	
12	4/25	API Coding for Cassandra	
13	4/27	NoSQL in the clude	
13	5/2	Map-Reduce, Hadoop Hive	
14	5/4	Presentation	
14	5/9	Presentation	
15	5/11	Presentation	
15	5/16	Presentation	
Final			