San Jose State University
College of Science
Department of Computer Science
CS157C, NoSQL Database Systems, Sections 1 and 2, Spring 2019

Course and Contact Information

- Instructor: Suneuy 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 9:20 am - 10:20 am
- Class Days/Time/Classroom
    - Section 1 (Lecture): MW 10:30 - 11:45, MH225
    - Section 2 (Lecture): MW 12:00 - 13:15, MH225
- Course Prerequisites: CS157A (or a grade of C- or better)
- Course Web Site 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 Data Models: Key-Value, Wide-Column, Document, and Graph Stores. CAP Theorem. Distribution Models. Current NoSQL Databases: Configuration and Deployment, CRUD operations, Indexing, Replication, and Sharding. Public Data Sets. API Coding and Application Development. NoSQL in the Cloud. Team Project.

Course Learning Outcomes

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

- Know the main NoSQL data models: Key-value, column-family, document, and graph stores
- Perform comparative analysis on NoSQL data models and relational data model
- Understand data distribution methods: replication and sharding
- Understand master-slave and peer-to-peer replications
- Understand Brewer's CAP Theorem and its implications for NoSQL database systems
- Understand the essentials of NoSQL data management through the CRUD operations and the querying mechanisms
- Understand NoSQL database system components and their communication protocols for the read and write process
- Select an appropriate NoSQL database for the use case at hand and design applications to efficiently work with the chosen database

Course Topics

Topics	Weeks

Features of NoSQL Databases	0.5
NoSQL Data Models	1.0
NoSQL Distribution Models	0.5
Introduction to MongoDB	0.5
MongoDB CRUD operations	0.5
MongoDB Advanced Queries	0.5
MongoDB Indexes	0.5
MongoDB Replication	0.5
MongoDB Sharding	0.5
API coding with MongoDB	0.5
Introduction to Cassandra	0.5
Cassandra Query Language (CQL)	1
Cassandra Data Modeling	1.0
Cassandra Architecture	1.5
NoSQL in the cloud: Amazon AWS	0.5
Project proposal presentations	1.5
Paper presentations	2.5
Total	14

Note: Selection of specific NoSQL databases may vary, but should be chosen to compare and contrast data models (e.g., document vs. column-family store) and distribution models (e.g., master-slave vs. peer to peer distribution). For any chosen NoSQL databases, their configuration and deployment, CRUD operations, strategies of indexing, replication, and sharding are expected to be taught.

Required Texts/Readings

- Textbook: None required
- References
  - 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 additional references will be provided per topic as needed.

Course Requirements and Assignments

- Programming Assignments: 3-4 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 assignme nt 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 responsibility and contribution of every team member must be precisely documented in a peer evaluation form.
- Software (Students are responsible to set up and deploy required software products. The instructor may not involve with any troubleshooting.)
  - MongoDB
  - Cassandra
  - Oracle Virtual Box
  - Linux (Ubuntu, CentOS, etc)
  - Programming Language: Java and/or Python
- Success in this course is based on the expectation that students will spend, for each unit of credit, a minimum of 45 hours over the length of the course (normally three hours per unit per week) for instruction, preparation/studying, or course related activities, including but not limited to internships, labs, and clinical practica. Other course structures will have equivalent workload expectations as described in the syllabus.

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: See the schedule below.
  - Final Exam: See the schedule below.
- 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: 30%

    - Project: 15%
    - 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

- Policy on Academic Integrity
    - Any cheating on an exam will result in a grade of F in the class.
    - If duplicate programs are found, both the provider and the copier will receive 0 point on the assignment. A second offense results in a grade of F in the class.
    - Any incident of academic dishonesty will be reported to University for disciplinary action.
- Attendance: University policy F15-12 at http://www.sjsu.edu/senate/docs/F15-12.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."
- 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:
    - "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 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 at http://www.sjsu.edu/gup/syllabusinfo/"

CS157C: NoSQL Database Systems, Spring 2018: Semester Schedule

Subject to change with fair notice.

Week	Dates	Topics	Assignments
1	1/28	Course Orientation	
1	1/30	Features of NoSQL Databases	
2	2/4	NoSQL Data Models	
2	2/6	NoSQL Data Models	
3	2/11	NoSQL Distribution Models	
3	2/13	Introduction to MongoDB	

4	2/18	MongoDB CRUD operations	
4	2/20	MongoDB Advanced Queries	
5	2/25	MongoDB Indexes	
5	2/27	MongoDB Replication	
6	3/4	MongoDB Sharding	
6	3/6	API Coding with MongoDB	
7	3/11	Project Proposal Presentation	
7	3/13	Project Proposal Presentation	
8	3/18	Project Proposal Presentation	
8	3/20	Midterm (NoSQL in the cloud)	
9	3/25	Introduction to Cassandra	
9	3/27	Cassandra Query Language (CQL)	
	4/1	Spring Recess	
	4/3	Spring Recess	
10	4/8	Cassandra Query Language (CQL)	
10	4/10	Cassandra Data Modeling	
11	4/15	Cassandra Data Modeling	
11	4/17	Cassandra Architecture	
12	4/22	Cassandra Architecture	
12	4/24	Cassandra Architecture	
13	4/29	Presentation	
13	5/1	Presentation	
14	5/6	Presentation	
14	5/8	Presentation	
15	5/13	Presentation	
Final		Section 1: Wednesday, May 15 9:45 - 12:00 Section 2: Friday, May 17 9:45 - 12:00	