San Jose State University
College of Science
Department of Computer Science
CS157C, NoSQL Database Systems, Sections 1 and 2, Spring 2021

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: TR 10:30 am - 11:30 am at Zoom: https://sjsu.zoom.us/j/89084181214
- Class Days/Time/Classroom
  - Section 1 (Lecture): MW 9:00 am - 10:15 am at Zoom: Register in advance for this meeting: https://sjsu.zoom.us/meeting/register/tZYlcuqqrDIiH9BSKjyRiDNGZ8dk3w-fBlLQ
  - Section 2 (Lecture): MW 10:30 am - 11:45 am at Zoom: Register in advance for this meeting: https://sjsu.zoom.us/meeting/register/tZYrdeurpzsjH9yr7nGDLLsKGcGwxahT3BMg
- 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
Fundamenatals of NoSQL (NoSQL Features, Data Models, and Distributoin Models)	1.5
Introduction to MongoDB	1
MongoDB CRUD operations and Advanced Queries	2
MongoDB Replication	0.5
MongoDB Sharding	1
MongoDB Indexes	1
Introduction to Cassandra	0.5
Cassandra Query Language (CQL)	1
Cassandra Data Modeling	0.5
Cassandra Architecture	2
Neo4j	1
Project Presentations	2
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 (available online at SJSU library)
    - NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence by Parmod J. Sadalage and Martin Fowler
    - MongoDB: The Definitive Guide: Powerful and Scalable Data Storage, 3nd Edition by Kristina Chodorow, December 2020
    - The Definitive Guide to MongoDB: A Complete Guide to Dealing with Big Data using MongoDB, 3rd Edition by David Hows, Peter Membrey, Eelco Plugge and Tim Hawkins, December, 2015
    - Mastering Apache Cassandra 3.x, 3rd Edition by Nishant Neeraj, Tejaswi Malepati and Aaron Ploetz, October 2018
    - Cassandra: The Definitive Guide: Distributed Data at Web Scale by Jeff Carpenter and Eben Hewitt, July 2016
    - Seven Databases in Seven Weeks: A Guide to Modern Databases and the NoSQL Movement, 2nd Edition by Luc Perkins, Eric Redmond, and Jim Wilson, April 2018
    - Learning Neo4j 3.x, 2nd Edition by Jerome Baton and Rik Van Bruggen, October 2017
    - Other readings: A list of additional references will be provided per topic as needed.

Course Requirements and Assignments

- Assignments: 4-5 individual 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 25 minute-long project presentation per team.
    - The final result of the project will be submitted through the project submission link on the course web site.

- 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 responsibility and contribution of every team member must be precisely documented in a peer evaluation form.
- Software (Students are responsible for setting up and deploy required software products. The instructor may not involve with troubleshooting.)
  - MongoDB
  - Cassandra
  - Neo4j
  - Docker
  - GIT
  - Linux (Ubuntu)
  - 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 (Take-Home Exam): See the schedule below.
  - Final Exam (Traditional 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 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: 25%
  - Midterm: 20%

- Final Exam: 35%
- Project: 15%
- Project Presentation: 3% (earn credit as a speaker)
- Participation: 2% (earn credit by participating in project presentation and/or guest lecture as an audience.)

- 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.

Technology Requirements

Students are required to have an electronic device (laptop, desktop or tablet) with a camera and built in microphone. SJSU has a free equipment loan (https://www.sjsu.edu/learnanywhere/equipment/index.php) program available for students.

Students are responsible for ensuring that they have access to reliable Wi-Fi during tests. If students are unable to have reliable Wi-Fi, they must inform the instructor, as soon as possible or at the latest one week before the test date. See Learn Anywhere website (https://www.sjsu.edu/learnanywhere/equipment/index.php) for current Wi-Fi options on campus.

Recording Zoom Classes

This course or portions of this course (i.e., lectures, discussions, student presentations) will be recorded for instructional or educational purposes. The recordings will only be shared with students enrolled in the class through the course web site. If you would prefer to remain anonymous during these recordings, then please speak with the instructor about possible accommodations (e.g., temporarily turning off identifying information from the Zoom session, including student name and picture, prior to recording). Students are not allowed to record my Zoom classes.

Online Exams

Proctoring Software and Exams

Exams will be proctored in this course through Respondus Monitor and LockDown Browser. Please note it is the instructor's discretion to determine the method of proctoring. If cheating is suspected the proctored videos may be used for further inspection and may become part of the student's disciplinary record. Note that the proctoring software does not determine whether academic misconduct occurred, but does determine whether something irregular occurred that may require further investigation. Students are encouraged to contact the instructor if unexpected interruptions (from a parent or roommate, for example) occur during an exam.

Testing Environment: Setup

- No earbuds, headphones, or headsets
- The environment is free of other people besides the student taking the test.
- No other browser or windows besides Canvas opened.
- A workplace that is clear of clutter (i.e., reference materials, notes, textbooks, cellphone, tablets, smart watches, monitors, keyboards, gaming consoles, etc.)
- Well-lit environment. Can see the students' eyes and their whole face. Avoid having backlight from a window or other light source opposite the camera.

Students must:

- Remain in the testing environment throughout the duration of the test.
- Keep full face in full view of the webcam

Technical difficulties

Internet connection issues: Canvas autosaves responses a few times per minute as long as there is an internet connection. If your internet connection is lost, Canvas will warn you but allow you to continue working on your exam. A brief loss of internet connection is unlikely to cause you to lose your work. However, a longer loss of connectivity or weak/unstable connection may jeopardize your exam.
Other technical difficulties: Immediately email the instructor a current copy of the state of your exam and explain the problem you are facing. Your instructor may not be able to respond immediately or provide technical support. However, the copy of your exam and email will provide a record of the situation.

Contact the SJSU technical support for Canvas:

```
Technical Support for Canvas Email: ecampus@sjsu.edu Phone: (408) 924--2337
https://www.sjsu.edu/ecampus/support/
```

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 2021: Semester Schedule

Subject to change with fair notice.

Week	Dates	Topics	Assignments

1	1/27	Course Orientation	
1	2/1	Fundamenatals of NoSQL	
2	2/3	Fundamenatals of NoSQL	
2	2/8	Fundamenatals of NoSQL	
3	2/10	Introduction to MongoDB	
3	2/15	Introduction to MongoDB	
4	2/17	MongoDB CRUD operations and Advanced Queries	
4	2/22	MongoDB CRUD operations and Advanced Queries	
5	2/24	MongoDB CRUD operations and Advanced Queries	
5	3/1	MongoDB CRUD operations and Advanced Queries	
6	3/3	MongoDB Replication	
6	3/8	MongoDB Sharding	
7	3/10	MongoDB Sharding	
7	3/15	MongoDB Indexes	
8	3/17	MongoDB Indexes	
8	3/22	Midterm (Take Home Exam - MongoDB in AWS cloud)	
9	3/24	Introduction to Cassandra	
	3/29	Spring Recess	
	3/31	Cesar Chavez Day (Observed - no class)	
9	4/5	Cassandra Query Lanagues (CQL)	
10	4/7	Cassandra Query Lanagues (CQL)	
10	4/12	Cassandra Data Modeling	
11	4/14	Cassandra Architecture	
11	4/19	Cassandra Architecture	
12	4/21	Cassandra Architecture	
12	4/26	Cassandra Architecture	
13	4/28	Neo4j	
13	5/3	Neo4j	
14	5/5	Project Presentation	
14	5/10	Project Presentation	
15	5/12	Project Presentation	
15	5/17	Project Presentation	
Final		Section 1: Wednesday, May 19 0715-0930 Section 2: Tuesday, May 25 0945-1200	