San José State University
College of Science / Department of Computer Science
NoSQL Database Systems, CS157C-03, Spring 2021

Course and Contact Information (Synchronous Online Course)

Instructor:	Dr. Mike Wu
Office Location:	MacQuarrie Hall 211(Online)
Email:	Ching-seh.Wu@sjsu.edu
Office Hours:	Tuesday 6pm-7pm & Friday 2pm – 3pm **(Please drop me an email with time info and subject.)**
Class Days/Time:	Tuesday and Thursday 4:30pm ~ 5:45pm
Classroom:	Online
Prerequisites:	CS 157A (with a grade of "C-" or better); Computer Science

Course Motivation

NoSQL (Non-SQL or Not-only-SQL) databases are increasing in popularity due to the growth of data as they can store non-relational data on a super large scale, and they can solve problems regular databases can't handle. They are widely used in Big Data operations. Their main advantage is the ability to handle large data sets effectively as well as scalability and flexibility issues for modern applications. There are different categories of NoSQL databases and they are used in social media such as with Facebook and search like Google and in other types of sectors like Health, Aviation, Education and other areas.

Catalog 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 (CLO)

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

Required Texts/Readings

Textbook: Not Required.

References

- NoSQL for Mere Mortals by Sullivan, 1st Edition. Addison-Wesley Professional, 2015 (SJSU library on-line access)
- SQL & NoSQL Databases: Models, Languages, Consistency Options and Architectures for Big Data Management by Meier & Kaufmann, 1st ed. Springer, 2019 (SJSU library on-line access)
- NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence by Parmod J. Sadalage and Martin Fowler (SJSU library on-line access)
- MongoDB: The Definitive Guide: Powerful and Scalable Data Storage, 3rd Edition by Kristina Chodorow, December 2019 (SJSU library on-line access)
- 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 (SJSU library on-line access)
- Mastering Apache Cassandra 3.x, 3rd Edition by Nishant Neeraj, Tejaswi Malepati and Aaron Ploetz, October 2018 (SJSU library on-line access)
- Cassandra: The Definitive Guide: Distributed Data at Web Scale Jul 22, 2016 by Jeff Carpenter and Eben Hewitt (SJSU library on-line access)
- Other readings: Additional references will be provided per topic as needed.

Representative Methods of Instruction:

- Lecture
- Directed Study
- Activity
- Discussion
- Other: Lecture will be used to introduce new topics. Instructor will model problem-solving techniques. Class will solve a problem together, each student contributing a potential "next step". Students will participate in short in-class projects (in instructor-organized small groups) to ensure that students experiment with the new topics in realistic problem settings. Instructor will invite questions AND ANSWERS from students, generating discussion about areas of misunderstanding. Instructor will create and manage an internet conference for discussion of course topics. Students will work in small groups to solve assignments.

Homework Assignments

You are expected to learn all of the material presented in the lectures. Written/programming assignments and project reports are also a requirement of the course. They must be turned in on time; **late homework/reports and email submission will NOT be accepted.**
Typical written assignments would include but not limited to:

- Designing a NoSQL database employing the NoSQL models
- Querying a database updating and deleting database content
- Writing applications that interact with NoSQL databases
- Employing XML and JSON to retrieve data

- Using NoSQL technologies to extract and manipulate web-based data
- Non-relational, distributed database design and creation using NoSQL web-based databases Write applications that use visualization and graphing to display data
- Use Big Data technologies such as Hadoop and MapReduce

HW assignments and project may involve installation/uninstallation of open source software.

Reading Assignments

Students will read assigned supplemental handouts.

Other Outside Assignments

Students will be required to watch video clips related to the lecture topics.

Interactive Pop Questions

Unannounced interactive pop questions may be given anytime during class. The purpose of pop questions to encourage you to learn, study and review the concepts and materials presented/discussed in the lecture. These will generally be problems covered in the previous lecture. You will be called to answer pop questions anytime during the online lecture. If your name is called and no response, 0 points will be recorded.

Midterm and Final Exams

Exams will consist of questions and problems aimed at assessing student mastery of course topics. Conceptual questions may be in the form of essay or multiple-choice format. Problem solving will require NoSQL programming code, data models, or similar output. All exams are closed books and notes.
If you are unable to attend any one of the exams, arrangements may be made only if you have a legitimate reason. You need to inform the instructor ahead of time and have written documentation available. If you are unable to attend the exam due to illness or emergency, you also need to inform your instructor **before the exam** and bring documentation afterwards to request a make-up exam, or the points for that exam will be allocated to other exams.

NoSQL Database Design, Implementation, and Deployment Project

The course achieves a balance between establishing a theoretical foundation and pragmatic applications of NoSQL in a real-world environment. A significant semester-long project reinforces lectures and is designed by applying Project Based Learning (PBL) derived from Google's software engineering best practices. In this team project, you will apply concepts presented in class and obtain practical, hands-on experience. A 3-member team will design, configure, implement, and deploy a small NoSQL database application. Team may choose any NoSQL database for the application that are appropriate in size and complexity. Appropriateness will be determined by the instructor. Students are responsible to set up and deploy required software products. The instructor may not involve with any troubleshooting.

By submitting/presenting a project, team members attest that they all participated in the conceptualization and accomplishment of the project. It is incumbent on team members to assure that **each team member MUST contribute in writing program code and documents** (Github will show each member's contribution to each file of code and document), no one on the team "free rides" through the project. If problems arise during the term, upon consultation with team members, the instructor will remove non-participating team members from their teams. Individuals removed from teams will not receive points on the team project.

Grading Information

Determination of Grades

The components of the final grade will be distributed as follows:

- **Class Participation: 10%** (Interactive pop questions, discussions, etc.)
- **Written & Programming Assignments: 30%** (6 Individual HWs)
- **Project: 20%** (Team with peer evaluations)
- **Midterm Exam: 20%**
- **Final exam: 20%** (Accumulative/Comprehensive)

Digit number grades will be assigned according to the following policy:

97 ~ 100 ---- A+
93 ~ 96 ---- A
90 ~ 92 ---- A-
87 ~ 89 ---- B+
83 ~ 86 ---- B
80 ~ 82 ---- B-
77 ~ 79 ---- C+
73 ~ 76 ---- C
70 ~ 72 ---- C-
67 ~ 69 ---- D+
63 ~ 66 ---- D
60 ~ 62 ---- D-
0 ~ 59 ---- F

Each assignment, project, and exam will be scored (given points) but not assigned a letter grade. Final individual class letter grades will be assigned based on the class curve.

Online Class Protocol and Other Notes

- **This course is defined by university as a synchronous online course very much like in-person class with a fixed weekly schedule so that there will be no lecture recordings.**
- **Absences in attending anyone of the first two lectures will be instructor-dropped out from the class.**
- Students are required to have an electronic device (laptop, desktop or tablet) with a camera and built-in microphone. 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.
- Even though this is an online course, every student must attend class and participate actively. Participation doesn't mean that you just log into your Zoom class. You must sit in the front of your computer (**no cell phone**) without any other activities at the same time.
- Exams will be proctored in this course through Respondus Monitor (with eye-tracking) and LockDown Browser. A webcam during exams is required. 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.
- This course is defined as a synchronous online course. There will be NO Zoom lecture recordings for later review/study. Recording a lecture is prohibited.

Students are prohibited from recording class activities (including class lectures, office hours, advising sessions, etc.), distributing class recordings, or posting class recordings. Materials created by the instructor for the course (syllabi, lectures and lecture notes, presentations, etc.) are copyrighted by the instructor. This university policy (S12-7) is in place to protect the privacy of students in the course, as well as to maintain academic integrity through reducing the instances of cheating. **Students who record, distribute, or post these materials will be referred to the Student Conduct and Ethical Development office.** Unauthorized recording may violate university and state law. It is the responsibility of students that require special accommodations or assistive technology due to a disability to notify the instructor.

- You may be called in most class sessions for pop questions and to discuss material contained in lectures by using Random Roster Checker.
- **When emailing me, please always start your email subject line with "CS157C: XXXXX" to get my attention. (for example: CS157C:HW1 Question)**
- **Plagiarism/Cheating will not be tolerable: 'F' will be given to your FINAL COURSE GRADE and will be reported to the Department and the University. The F grade will not be automatically turned to NC grade. (Obtaining HW solutions from someone or giving/showing your HW solutions to someone is also treated as plagiarism/cheating.)**
- **Participation is crucial to perform well on pop questions, assignments and examinations.** Regular attendance is your responsibility. If you choose to miss classes, it is also your responsibility to make up all work missed.
- **Students are responsible for all materials distributed on Canvas and discussed in the class.**
- **I reserve the right to make announcements in class that may not appear on the class website/Canvas.**

University Policies

Attendance: 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.

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 are not allowed to publicly share or upload instructor generated material for this course such as exam questions, lecture notes, or homework solutions without instructor consent.**

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/" Make sure to review these policies and resources.

'

Tentative Course Schedule (This schedule is subject to change with fair notice.)

Week	Date	Topics, Readings, Assignments, Deadlines
1	1/28	Motivation, Course Introduction, Prerequisites Check
2	2/2	Topic I: **Introduction to NoSQL** (What is NoSQL, NoSQL Overview, NoSQL Database Environment, NoSQL Options)
2	2/4	Topic II: **When to use NoSQL** (Benefits to using NoSQL DB • Backend Management, Deployment, Front-End Development)
3	2/9	Topic II: **When to use NoSQL (continued)** (Open Source, Drawbacks to Using NoSQL DB, NoSQL vs. SQL)
3	2/11	Problem Solving and Hands-on Session
4	2/16	Topic III: **Introduction to NoSQL Development** (Schemaless Development, Data Models, Distribution Models, Consistency) **Project team formation, 3 students per team**
4	2/18	Topic III: **Introduction to NoSQL Development (Continued)** (Categories of NoSQL: Key-Value Stores, Wide-Column Family Stores, Document Databases, Graph Databases, Object-Oriented Databases, and Others, NoSQL Scalability, Searching) **Project Proposal Discussions** **GitHub account creation for team project due**
5	2/23	Problem Solving and Hands-on Session **Project Proposal & Requirements Document Due**
5	2/25	Topic IV: **Wide-Column NoSQL Databases** (Column Family, Key and Keyspace, Categories of NoSQL)
6	3/2	Topic IV: **Wide-Column NoSQL Databases (Continued)** (Example NoSQL Databases: Cassandra, Bigtable, MapR, and Others)
6	3/4	Problem Solving and Hands-on Session
7	3/9	Topic V: **Key-Value NoSQL Databases** (Major keys, Minor keys, Values)
7	3/11	Topic V: **Key-Value NoSQL Databases (Continued)** (Example NoSQL Databases: Oracle NoSQL Database, Redis, and Others)
8	3/16	Problem Solving and Hands-on Session
8	3/18	Topic VI: **Document NoSQL Databases** (Attributes, Metadata, Formats, XML, JSON and BSON)
9	3/23	Topic VI: **Document NoSQL Databases (Continued)** (Example NoSQL Databases: ElasticSearch, Mongo DB, Couch DB and Others)
9	3/25	Problem Solving and Hands-on Session **(Project Data Model & NoSQL DB Design Document Due)**
	3/30	Spring Recess 3/29 ~ 4/2 (No Class)
	4/1	Spring Recess 3/29 ~ 4/2 (No Class)
10	4/6	Topic VII: **Graph NoSQL Databases** (Building Graph Model, Edges, Nodes, Relationships) **First Project Code Review (Revision Code due ion Github)**
10	4/8	Topic VII: **Graph NoSQL Databases (Continued)** (Example NoSQL Databases: Neo4J, InfoGrid)
11	4/13	Problem Solving and Hands-on Session

Week	Date	Topics, Readings, Assignments, Deadlines
11	4/15	Topic VII: **Graph NoSQL Databases (Continued)** (Example NoSQL Databases: GraphBase and Others) Second Project Code Review (Revision code due in Github)
12	4/20	Problem Solving and Hands-on Session
12	4/22	Topic VIII: **Cloud Computing with NoSQL Database** (Big Data, Remote Searches)
13	4/27	Topic VIII: **Cloud Computing with NoSQL Database (Continued)** (Hadoop, MapReduce)
13	4/29	Problem Solving and Hands-on Session Third Project Code Review (Revision code due in Github)
14	5/4	Topic VIII: **Cloud Computing with NoSQL Database (Continued)** (Rest, AWS)
14	5/6	Problem Solving and Hands-on Session
15	5/11	Final Project Presentation & Demo
15	5/13	Final Project Presentation & Demo Final Review Final Project Code Review (final revision code due in Github)
Final Exam	5/21	Friday 2:45pm – 5:00pm (Final Project Report Due) (Project Peer Evaluation Due)