**DSO 553: NoSQL Databases in Big Data (1.5 credits)**

**Location:** to be added

**Instructor:** Joydeep Banerjee

**Contact Info:**joydeep@marshall.usc.edu

**Course Description**

NoSQL databases or “not only SQL databases” play a pivotal role in today’s world of Big Data. This course provides the student with an understanding of how and why they happened and the hard problems it solved. It will introduce the idea of structured, semi structured and unstructured data. Given the various data sources and structures of data that companies are dealing with today, this course provides the framework and understanding to assess the NoSQL ecosystem holistically and then engage with each at a deep level - adding to their knowledge of database design using relational techniques only. This is a deeply practical course dealing with what is happening in the industry today. It will also equip the student to gain hands on knowledge on two of the most widely used NoSQL databases – Document Database (MongoDB) and Graph database (Neo4j).

**Learning Objectives**

Upon successful completion of this course, student should be able to

1. Explain what is Big Data and what are its challenges
2. Describe the Big Data analytics, storage and computation landscape with introduction to terms like data lake and data warehouse
3. Explain the need for NoSQL database and different types of them and when to use one
4. Understand MongoDB, how to model, query and optimize MongoDB applications and compare them to RDBMS
5. Understand Neo4j, how to model, query and optimize Neo4j applications and compare them to RDBMS
6. Explain how these participate in the AI/ML ecosystem.

**Prerequisite(s):** DSO 552

**Course Notes:** Grades will be posted on Blackboard.

**Technological Proficiency and Hardware/Software Required**

1. Must have access to a laptop and internet for class and homework.
2. Connect to the slack channel (uscmsbclass.slack.com) for the class spring23-dso-553-16292 | spring23-dso-553-16294 | spring23-dso-553-16296 as the case may be. This can be used as “virtual office hours”.

**Required Readings and Supplementary Materials**

Readings are available via USC library and on Blackboard

*Also see*:

-NoSQL for Mere Mortals (optional) - via USC library

-Next Generation Databases: NoSQL, NewSQL, and Big Data (optional) - via USC library

-MongoDB: The Definitive Guide, 3rd Edition (optional) - via USC library

-Seven NoSQL Databases in a Week (optional) - via USC library

-[Other books here](https://learning.oreilly.com/playlists/5eac7569-0fef-47e0-b5fa-6807a8de3ae3) must sign in via USC library for O’Reilly

*Journals, Articles, Videos, Links*:  
[An Introduction to NoSQL Database Design, Technical Note](https://hbsp.harvard.edu/import/892448) (HBR)

[Big Data Architectures](https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/) (Azure documentation)

[Introduction to NoSQL Databases by Martin Fowler](https://www.youtube.com/watch?v=qI_g07C_Q5I)

**Grading Breakdown:**Grades will be calculated based on the following components. Please note that exceptional work or effort by a student will be recognized and rewarded. Similarly, plagiarism in written assignments will be penalized according to USC rules and guidelines.

|  |  |
| --- | --- |
| **Category** | **% of Overall Grade** |
|  |  |
| *Class Participation* | 10 |
| *Homework (6)* | 60 |
| *Finals* | 30 |
| *Total* | 100 |

**Class Participation**

Class sessions will provide helpful information for learning the topics covered in the course and working on the Project. Student participation in class discussions is crucial because it introduces alternative viewpoints and helps clarify concepts for the class. Cold calling may take place to encourage active participation and to gain multiple perspectives and points of view, thus lending itself to the richness of the learning experience. Participation grades will be based on the quality of a student's contribution to the lectures. The final participation grade will be determined solely at the discretion of the instructor. 10%

**Assignment Submission Policy**

1. Late assignments will receive a 1% deduction for every day after the assigned due date on Blackboard .
2. If you have difficulties completing the assignment, please let me know **AT LEAST 48 HOURS (2 DAYS)** ahead of the assignment due date on Blackboard.

**Grading Timeline**

Any grading disputes should be brought to my or the grader’s attention **within 3 days** of grades returned/posted.

**Course Schedule** (weekly)

|  |  |  |
| --- | --- | --- |
| **Class** | **Topic** | **Tasks/Homework** |
| Week-1  (6th March) | *Introduction to Big Data, NoSQL. Best use cases for Relational databases and different types of NoSQL databases.* | Set up MongoDB in the cloud and connect to it |
| Week 13th March | *Spring Break* |  |
| Week-2  (20th March) | *Deep Dive into MongoDB: Creating MongoDB collections, indexes and querying. Comparison with RDBMS sibling.* | HW-1 Due before class |
| Week-3  (27th March) | *Advanced MongoDB Concepts: Aggregate MongoDB pipeline. Comparison with RDBMS sibling.* | HW-2 Due before class |
| Week-4  (3rd April) | *Review MongoDB, MongoDB Use cases and introduction to Graphs and concepts of Nodes and Edges. Brief review of Linked Data and Knowledge graphs in public domain* | Set up Neo4j sandbox & HW-3 Due before class |
| Week-5  (10th April) | *Deep Dive into Neo4j.Creating Neo4j applications and querying and indexing.* | HW-4 Due before class |
| Week-6  (17th April) | *Advanced Neo4j concepts and comparison with recursive CTOs (RDBMS). Look at Graph Use cases.* | HW-5 Due before class |
| Week-7  (24th April) | *Quick look at other types of NoSQL databases – Key-Value, Columnar store. Quick introduction to other types of widely used data bases for Search and Time Series.* | HW-6 Due before class |
|  | ***And Finals*** |  |

**Statement on Academic Conduct and Support Systems**

**Academic Conduct:**

Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in *SCampus* in Part B, Section 11, “Behavior Violating University Standards”<https://policy.usc.edu/scampus-part-b/>. Other forms of academic dishonesty are equally unacceptable. See additional information in *SCampus* and university policies on scientific misconduct, [http://policy.usc.edu/scientific-misconduct](http://policy.usc.edu/scientific-misconduct/).

**Support Systems:**

Students whose primary language is not English should check with the American Language Institute http://dornsife.usc.edu/ali, which sponsors courses and workshops specifically for international graduate students.

*Student Counseling Services (SCS) - (213) 740-7711 – 24/7 on call*

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.<https://engemannshc.usc.edu/counseling/>

*National Suicide Prevention Lifeline - 1-800-273-8255*

Provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week. [http://www.suicidepreventionlifeline.org](http://www.suicidepreventionlifeline.org/)

*Relationship & Sexual Violence Prevention Services (RSVP) - (213) 740-4900 - 24/7 on call*

Free and confidential therapy services, workshops, and training for situations related to gender-based harm.<https://engemannshc.usc.edu/rsvp/>

*Sexual Assault Resource Center*

For more information about how to get help or help a survivor, rights, reporting options, and additional resources, visit the website:<http://sarc.usc.edu/>

*Office of Equity and Diversity (OED)/Title IX compliance – (213) 740-5086*

Works with faculty, staff, visitors, applicants, and students around issues of protected class.<https://equity.usc.edu/>

*Bias Assessment Response and Support*

Incidents of bias, hate crimes and microaggressions need to be reported allowing for appropriate investigation and response.<https://studentaffairs.usc.edu/bias-assessment-response-support/>

*Student Support & Advocacy – (213) 821-4710*

Assists students and families in resolving complex issues adversely affecting their success as a student EX: personal, financial, and academic.<https://studentaffairs.usc.edu/ssa/>

*Diversity at USC*

Tabs for Events, Programs and Training, Task Force (including representatives for each school), Chronology, Participate, Resources for Students. [*https://diversity.usc.edu/*](https://diversity.usc.edu/)

*USC Emergency Information*

Provides safety and other updates, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible. [emergency.usc.edu](http://emergency.usc.edu)

*USC Department of Public Safety – UPC: (213) 740-4321 – HSC: (323) 442-1000 – 24-hour emergency or to report a crime.* Provides overall safety to USC community. [dps.usc.edu](http://dps.usc.edu/)

**Students with Disabilities:**

USC is committed to making reasonable accommodations to assist individuals with disabilities in reaching their academic potential. If you have a disability which may impact your performance, attendance, or grades in this course and require accommodations, you must first register with the Office of Disability Services and Programs ([www.usc.edu/disability](http://www.usc.edu/disability)). DSP provides certification for students with disabilities and helps arrange the relevant accommodations. Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me as early in the semester as possible. DSP is located in GFS (Grace Ford Salvatori Hall) 120 and is open 8:30 a.m.–5:00 p.m., Monday through Friday. The phone number for DSP is (213) 740-0776. Email: ability@usc.edu.