|  |  |
| --- | --- |
| Jongwook Woo, Ph.D.  Office: Simpson Tower, 604  Telephone: Direct line:(323) 343- 2916  CIS Office: (323) 343‑2911  Email: [jwoo5@calstatela.edu](mailto:jwoo5@calstatela.edu) | **Office Hours (Zoom):**  Monday: 3:45 – 4:30 pm  Wednesday: 4:30 – 5:45 pm  **NOTE**: Zoom link should be given at the first class |

CIS5200 Systems Analysis and Design

PREREQUISITES:

1. Mastery over MS-Windows File Management (Windows Explorer) facilities.

**COURSE INFORMATION and DESCRIPTION**

* Life cycle of systems development; data modeling; process modeling; enterprise system integration, configuration, installation and maintenance; a case-based approach with hands-on experience in designing information systems, especially using Cloud Computing and Big Data Systems.

**COURSE GOALS/LEARNING OBJECTIVES:**

* Students will explain effective behavioral and technical techniques in Cloud Computing on IT and Big Data
* Students will deliver the genesis of Big Data Systems and explain its importance to improving the success of information technology projects
* Students will explain the knowledge of Big Data in industry and its Architecture
* Students will conduct data analysis, modeling and visualization in Big Data systems

**Textbook:**

Instructor’s lecture and lab materials will be posted at a web when the class starts.

**Course delivery structure:**

* Remote Lecture: Online
* Online lecture is either Synchronous or Asynchronous lecture and lab

**CLASSROOM EXPECTATIONS:**

Students are expected to attend every class session. Since Cloud computing and Big Data concepts are presented during class time, class attendance is essential for successful completion of assignments and tests. Since a large part of the course involves work on cloud computing and cluster, it is essential that you utilize the time in class for discussion and exercises on the computer. If you don’t, you would not catch up the class and then you would fail. If attendance is not possible for one of the class meetings, please contact the instructor beforehand.

Students are expected to use the equipment of the computer labs at CSULA.

Students who work off-campus or who take too many courses may want to drop this course if you think you cannot focus on assignments and cannot attend.

**ASSIGNMENTS and Grading Policy:**

Grading is based on instructor’s overall judgment of student’s performance. Student’s performance is measured based on the following elements:

* Class Activities (Pop quizzes, Attendance, Participation in Class):                 10%
* 2 -3 Individual Homework (Questions or Project Assignment)                     15%
* Labs:                                                          20%
* Midterm Exam 1:                                      7%
* Midterm Exam 2:             8%
* Final Team Project Presentation:                                               25%
* Final Term Paper:                                                           15%
* Total 100%

GRADE

At the end of the quarter, you will have a score out of 100 percent. This score will be used in a class curve to arrive at a letter grade. I guarantee that >= 90 will be some kind of A (A- or A), >= 80 will at least be some kind of B (B-, B, B+), >= 70 will at least be some kind of C (C-, C, C+), and that >=60 will be at least some kind of D (D-, D, D+).

**Academic dishonesty**

Academic dishonesty includes (but not limited to) the following:

1. Giving or receiving information during an exam.
2. Giving or receiving (share) solution for an individual assignment.

(The instructor can find the copies and the violator will have F on the assignment or the course)

1. Unauthorized or malicious use of computing facilities.
2. Deception or misrepresentation in a student's dealing with the instructor.
3. Inappropriate collaboration on or coping of homework assignments.

(Students are encouraged to discuss the readings with one another, even when the discussion relates to assignments. As long as the purpose of discussion is to help the student's understanding of the material, and not to reduce or share the work, such discussion will not be deemed inappropriate.)

1. Plagiarism, the submission of material authored by another person but represented as the students’ own work. It does not matter whether the original work author gave permission.
2. Any violation of academic integrity standards described in the student conduct code. Students are expected to be familiar with these standards.

**Class Communication**

1. **Discussion module of Canvas**

Debugging a program is often an iterative process of trail and error. Therefore, requests for debugging your program via Discussion Module of canvas can be actively answered.

1. **Use of email:**

Email will be used only for **short messages** and sending attachments of **less than one** Mega Byte.

Debugging a program is often an iterative process of trail and error. Therefore, requests for debugging your program via **email will not be responded but via Discussion.**

**A Tentative Course Schedule is as follows:**

## Jongwook Woo, CIS 5200, A Tentative Course Schedule

(Syllabus may be subject to change; mostly follow the lecture notes schedule at the Canvas of the course)

|  |  |
| --- | --- |
|  | Reading Assignment / Homework |
| Week 1**:** Course Overview  **Lecture 1** AnIntroduction to Big Data and Cloud Computing Systems | Reading Instructor’s material about the systems of Big Data and Cloud Computing |
| Week 2: Lecture 2 Big Data system development  * 1. Introduction to Hadoop   2. Motivation for Hadoop   **Lab 1:** set up cloud computing accounts such as Oracle Big Data Compute Edition | Reading Instructor’s material about Hadoop Homework 1 |
| Week 3: **Lecture 3** Basic Concepts: HDFS, MapReduce, Hive  **Lab 2**: HDFS and Hive in Oracle Big Data | Reading Instructor’s material about HDFS, MR, Hive Team Build for term project: email the preferred team members |
| Week 4: **Lecture 4** Initiating Phase: Basic Concepts: MR cont’d; Cluster; Ecosystems, Hive  **Lab 3**: Hive Web Log Analysis in Oracle Big Data | Reading Instructor’s material about MR, Cluster, Ecosystems, Hive  **Team built and choose topics for the team project** |
| Week 5: **Lecture 5** Hive Data Processing (Join, Union)  **Lab 4:** IoT Sensor Data Analysis using Hive in Oracle Big Data | Reading Instructor’s material about Join in HiveHomework 2 |
| Week 6: **Lecture 6** Text Analysis in Hive  **Lab 5:** Twitter Data Text Analysis using Hive in Oracle Big Data | Reading Instructor’s material about Hive Text Analysis |
| Week 7: Midterm Exam 1 Multiple Choice and Hands-on Exercises | Midterm Exam **(May 7: covers lecture 1 – lecture 6)** |
| Week 8: **Lecture 7** Sqoop and Hive  **Lab 6:** Movie Data Analysis using Sqoop and MySQL DB | Reading Instructor’s material about Sqoop and HiveHomework 3 |
| Week 9: Lecture 8 NGram, Text Processing Functions in Hive **Lab 7:** NGram Sentiment Text analysis of Twitter social media data | Reading Instructor’s material about NGram and Functions of Hive |
| Week 10: Lecture 9 Pig Fundamentals **Lab 8:** Pig Fundamentals in Oracle Big Data | Reading Instructor’s material about Pig and PigLatin |
| Week 11: Lecture 10 Pig Data Filtering and ProcessLab 9: Pig Data Process in Oracle Big Data | Reading Instructor’s material about Data Filtering and Process in Pig |
| Week 12: Lecture 11 Pig and Hive using HCatalogLab 10: HCatalog with IoT data of TruckEvent in Oracle Big Data | Reading Instructor’s material about Pig and Hive using HCatalog |
| Week 13: Lecture 12 Pig UDF and Streaming DataLab 11: Pig ETL Processing and visualization using Tableau in Oracle Big Data | Reading Instructor’s material about UDF and Streaming Data |
| Week 14: Midterm Exam 2 | Hands-on Exercise with Hive and Pig |
| Week 15: Term Project Presentation | Present Group Term Project: Topics of Big Data Analysis and Visualization using Cloud Computing |
| Week 16: Final Exam (NA) | No Final Exam  Submit Term Paper and Term Project Tutorial |