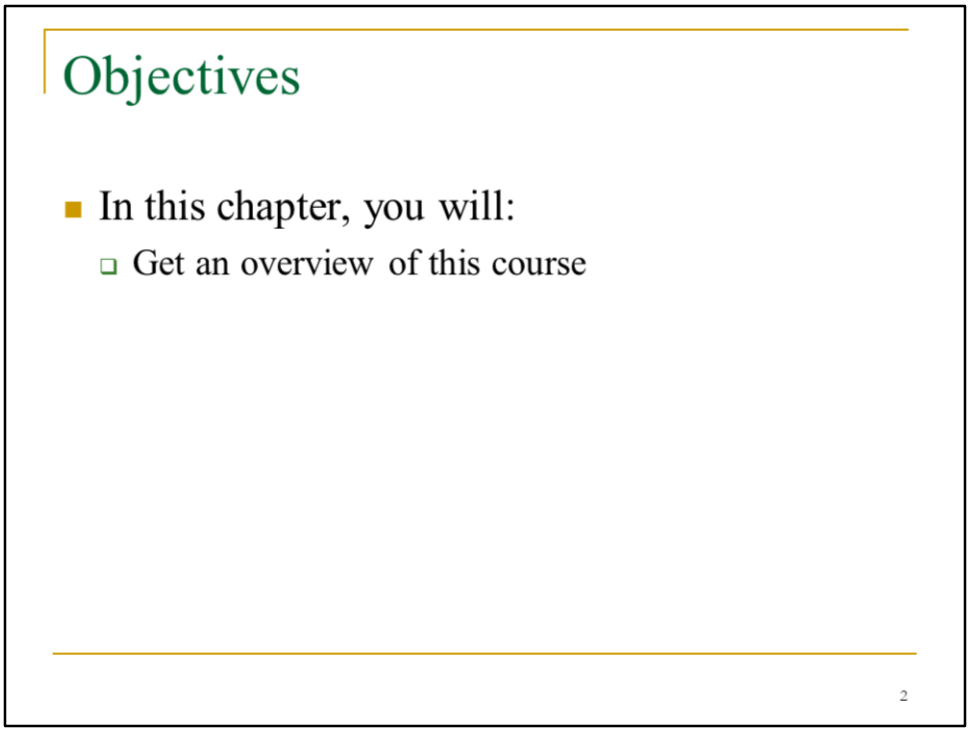
Big Data Analytics

Chapter 1: Course Introduction

1

Hello, everyone! Welcome to the Big Data Analytics class. In this chapter, we will talk about the course introduction.



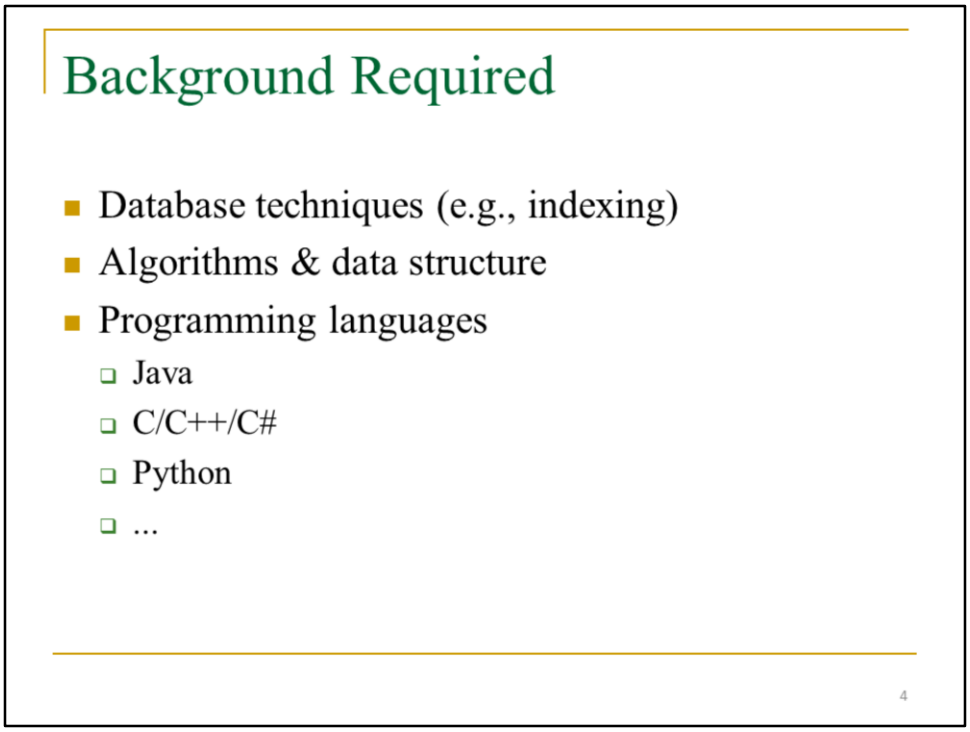
We will briefly review the content of this course.

Big Data Analytics

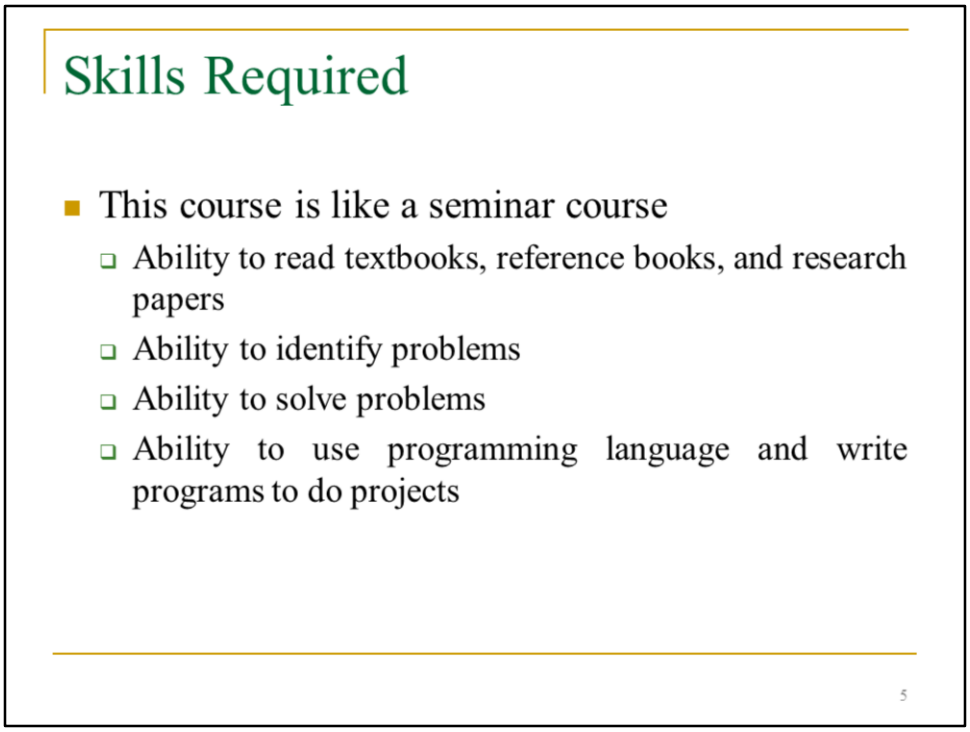
- Instructor: Dr. Xiang Lian
  - Homepage: http://www.cs.kent.edu/~xlian/
  - Office: MSB 264, Computer Science Department, Kent State University
  - Email: xlian@kent.edu
- Course website:
  - Blackboard: https://learn.kent.edu/

3

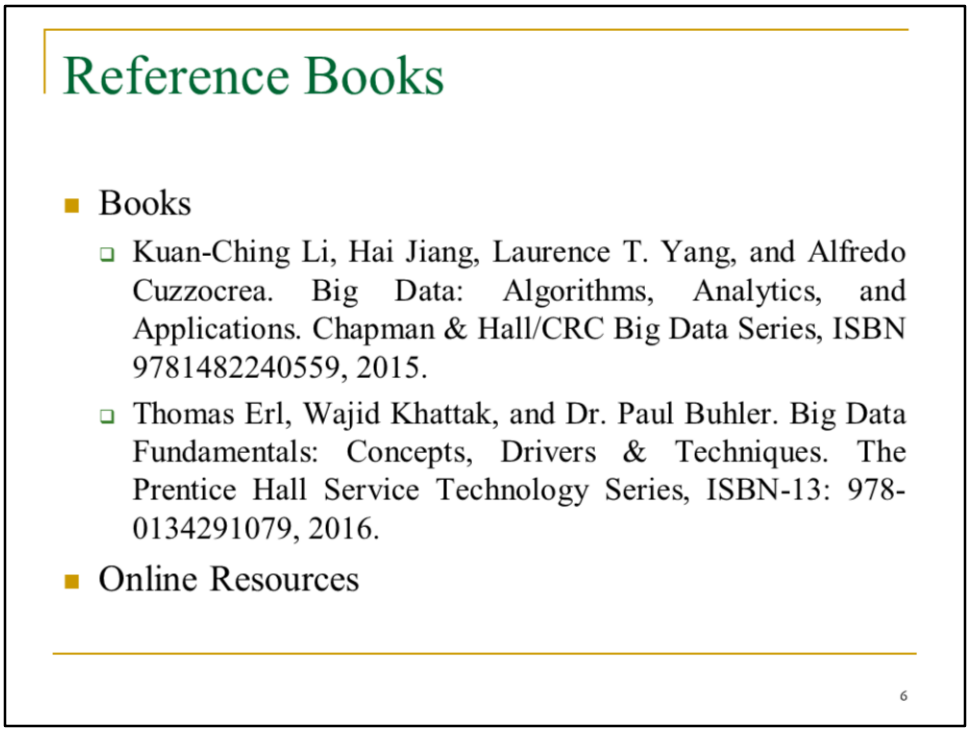
I am Dr. Xiang Lian, the instructor of the Big Data Analytics class. Here are the link to my homepage, my office at the Computer Science Department of Kent State University, and my email address. If you have any questions about course materials or works, you are welcome to send me emails and I will try my best to answer your questions. Or you can schedule a face-to-face meeting with me. Our course materials or assignments will be released on the Blackboard, and you also need to submit solutions to assignments, survey, or project report to the Blackboard.



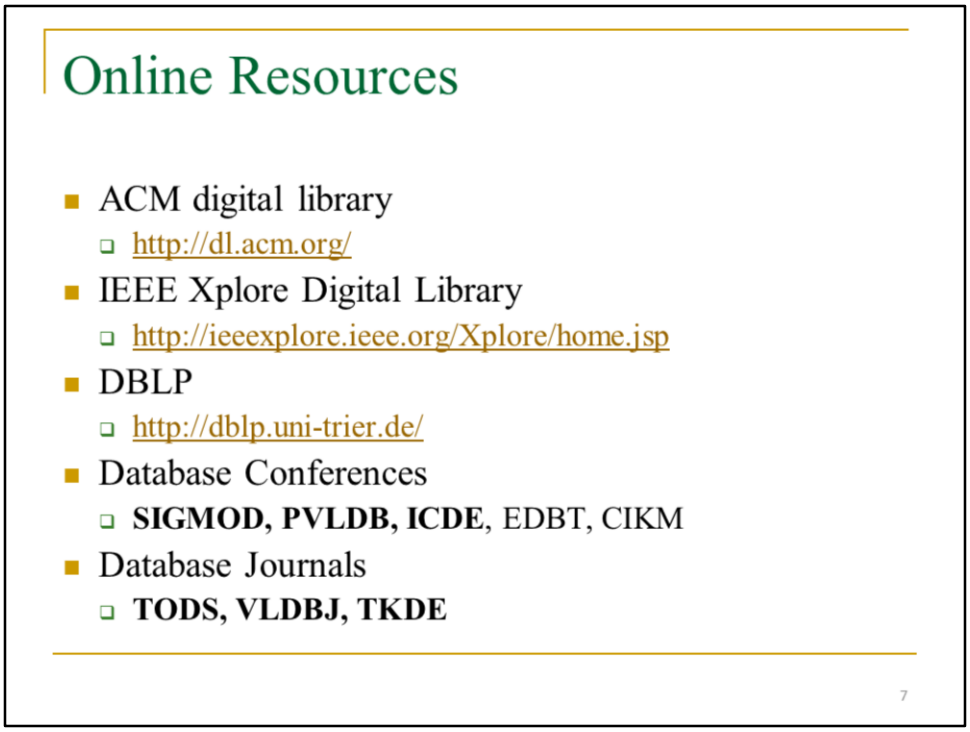
To take this course, you need some background such as the database techniques like indexing, algorithm and data structure, and programming skills. In this course, you need to implement a course project, by Java, C/C++, Python, or any other programming language.



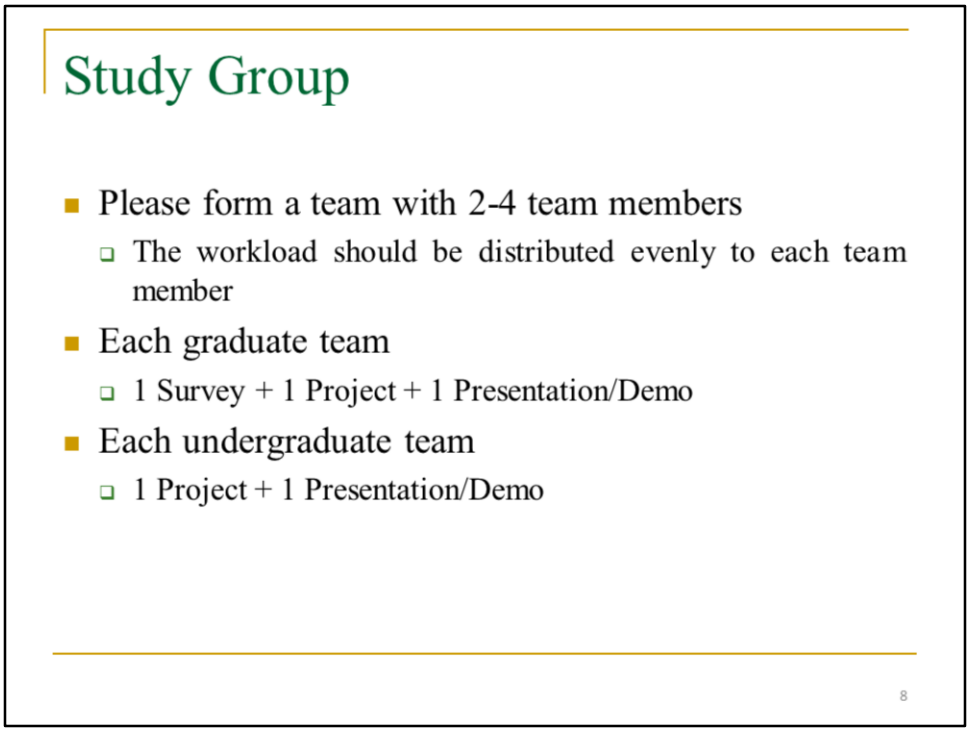
This course is like a seminar course. I will first give lectures about big-data-related topics, and meanwhile you need to read many course materials such as textbooks, reference books, and research papers. Therefore, this course requires your ability of reading course materials, identifying research problems, and finding solutions to problems.



There are no textbooks for this Big Data Analytics class. However, we do have some reference books about big data listed here. Also there are many online resources that you should use for the survey or project.

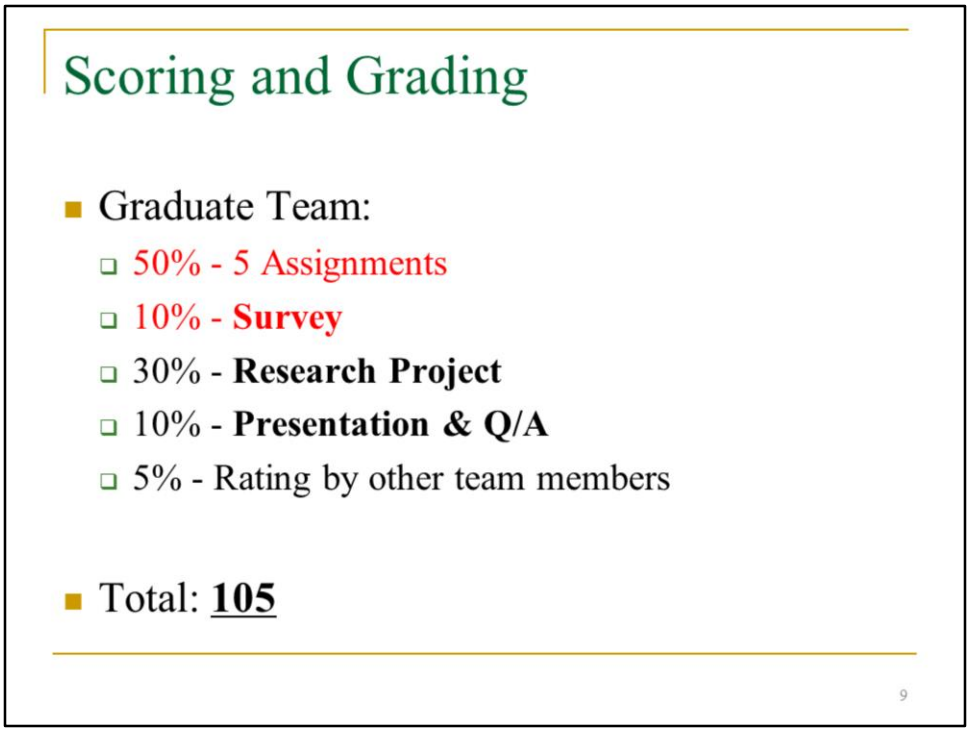


For example, you can find research papers through ACM digital library, IEEE Xplorer, and DBLP. Usually, when you do surveys and research projects, you may need to read some related works on research topics. You should search for papers from top database conferences such as SIGMOD, PVLDB, ICDE, EDBT, and CIKM, or top database journals such as TODS, VLDBJ, and TKDE.

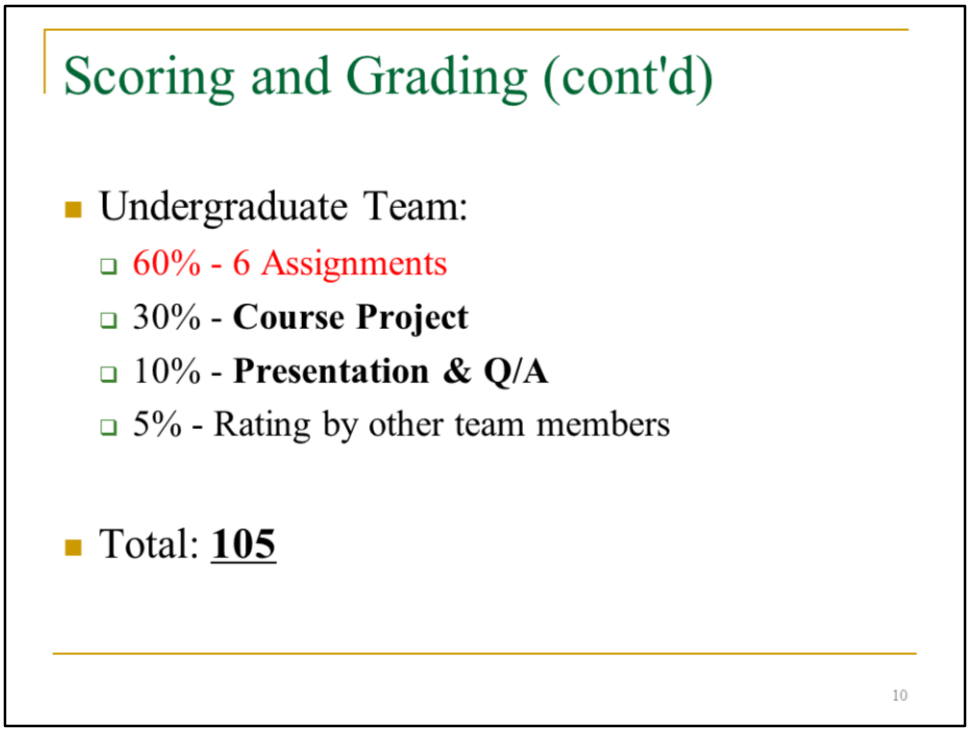


In this course, we will form study groups, each with 2-4 team members. All members in the same group must be either graduate or undergraduate students (i.e., cannot be a mix of graduate and undergraduate students).

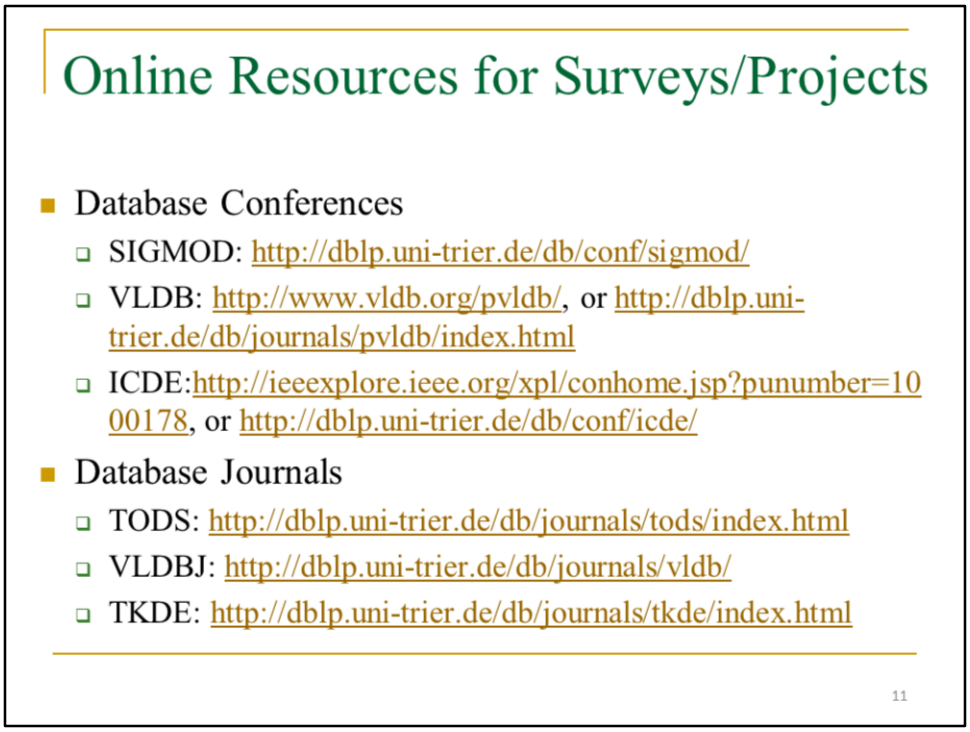
Throughout the semester, each graduate study group will do a survey, a course project, and a presentation/demo together; each undergraduate study group will do a course project, and a presentation/demo. Note that, the survey here is only for graduate students. You need to read a number of relevant research papers and write a summary of those papers. For the project, each group needs to either design new solutions to an existing problem, or propose solutions to a new problem. The presentation/demo is to present the solutions to your project problems. The workload of each group assignment should be distributed evenly to each team member.



Here is the scheme of the scoring and grading. In this semester, for graduate team, graduate students will do 5 individual assignments, each with 10 points, a group survey with 10 points, a group research project with 30 points, a group presentation in the form of the video (including Q/A) with 10 points, and ratings from other team members, which is 5 points. Note that, to prevent the case that some team members do not contribute much to group assignments, we introduce this peer evaluation method among team members. Please refer the grading scales to the course syllabus.



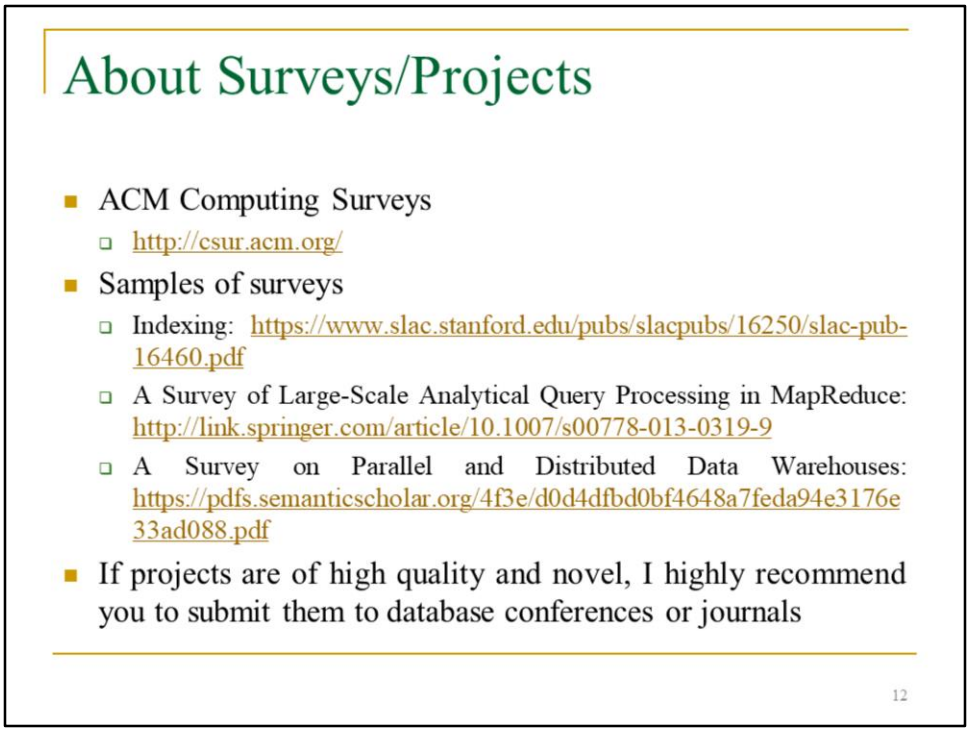
For undergraduate team, undergraduate students will do 6 individual assignments, each with 10 points, a group course project with 30 points, a group presentation in the form of the video (including Q/A) with 10 points, and ratings from other team members.



Here are links to tier-1 database conferences or journals.

For the survey, you need to go to each conference/journal website, and search for a list of related paper titles containing keywords (such as “big data”, “MapReduce”, “query”, etc.) in the most recent 3-5 years.

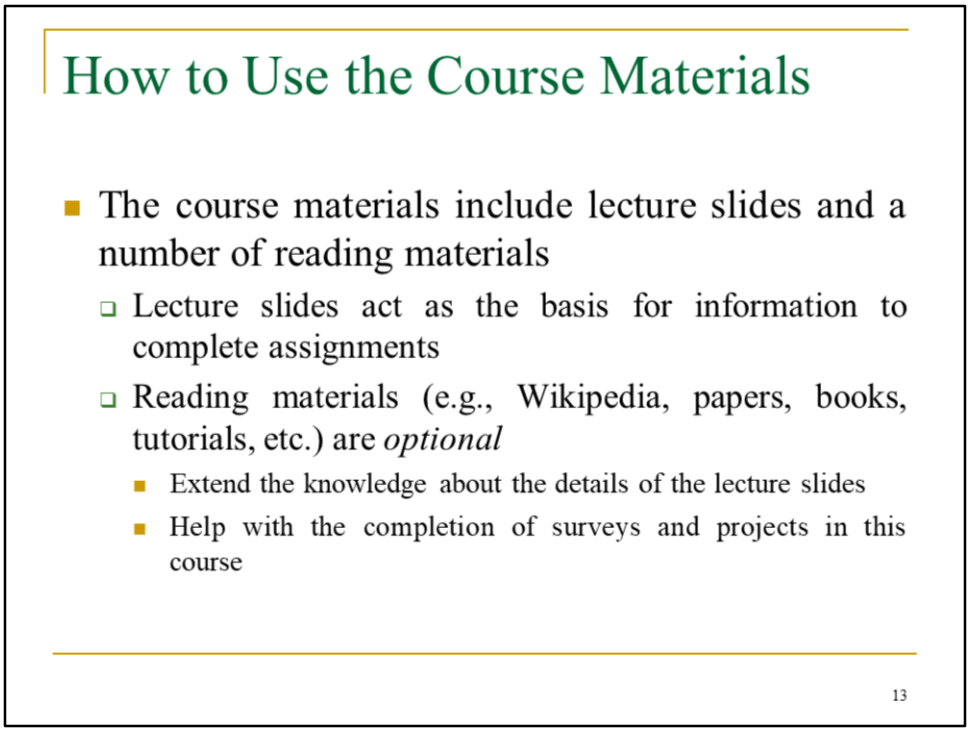
Similarly, for the project, you may also need to find previous solutions to an existing problem, and design a more efficient solution to an existing/new problem.



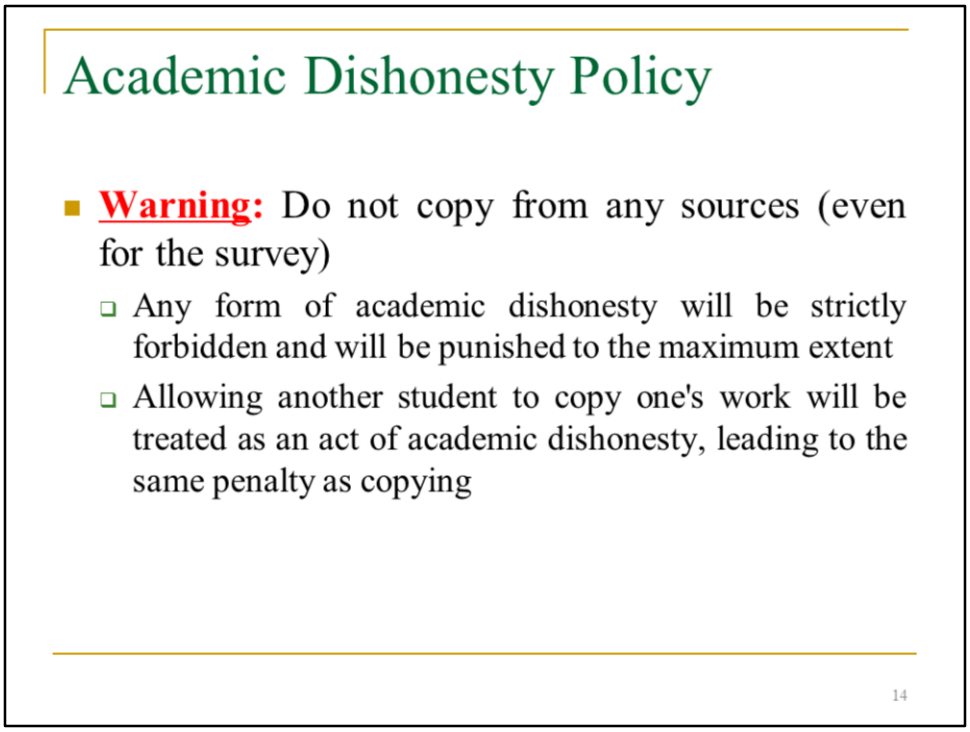
You can find many surveys from ACM Computing Surveys. This is a quick way to get familiar with a research area that you did not touch.

I also list some examples of surveys. You can check the details. They classify papers into different categories (or sub-categories), for example, AI-based and non-AI-based approaches, and summarize papers from each (sub-)category.

For the project, you first need to find a topic for the project, either a completely new problem or an existing problem but with new solution. If your project is of high quality and novel, I highly suggest you submit them to database conferences or journals.



This course provides a number of lecture slides and reading materials. Students are required to read and understand lecture slides, since they are important for completing the assignments. On the other hand, the reading materials such as Wikipedia, papers, books, video tutorials, and so on are optional. Nevertheless, they can help students extend their knowledge about lecture slides or topics for big data analytics, and help students finish surveys or projects in this course.



This course forbids any form of academic dishonesty. You should not copy from any other sources. Even for the survey conducted by graduate teams, you should use your own words to summarize the papers. Give appropriate citations of papers, figures, or tables that are not your contributions.

Note that, homework is not a group assignment. For the individual homeworks, you should do them independently. Please do not copy from other students or give copies to other students. In both cases, students will get zero point for copying assignments.