ACM Data Science Task Force Course Example

Fundamentals and Applications of Big Data
Beijing Institute of Technology, Beijing
Weixing Ji, Xin Sun, Yujin Gao

Knowledge Areas that contain competencies (knowledge, skills, and dispositions) covered in the course

Knowledge Area	Total Number of Contact Hours
Data Visualization	5
Data Structure	3
Data Analysis	7
Big Data Applications	4
Data Processing	13

Where does the course fit in your undergraduate Data Science curriculum?
This course is for non-information majors and is an elective course. It doesn't have pre-requisites or required following courses. The "Python Big Data Fundamentals" in the course mainly teaches the basic programming knowledge needed for big data analysis and design, including data types, data structures, data access, Part of the content such as data processing and display, and related online practice content, so you can learn this course without programming foundation. Male students have a higher proportion of students who take this course.

What is covered in the course?
(1) This course teaches the basic principles, related technologies and typical application cases of big data analysis. Learners can understand the construction methods and processes of big data application systems, cultivate the awareness of big data thinking, and master the basic technologies and methods of data analysis.

(2) This course uses the big data processing process as a clue to introduce data collection, data cleaning and processing, data analysis, open source big data computing platform, data visualization, etc. An introductory introduction to the commonly used big data programming language Python.

(3) The course introduces a large number of big data application cases, these cases come from different industries and fields, and provide students with a multi-faceted reference for different disciplines.

(4) The course designed an online big data practice system based on Huawei's public cloud. While learning theoretical knowledge, according to the experimental manual, students gain hands-on practice and experience of various algorithms and tools.

(5) The course introduces the needs of big data talents and common jobs, and provides reference and help for learners in their own career planning.

What is the format of the course?
The total class hours of this course are about 32 hours. It is an introductory course designed to enhance students' understanding of big data majors. In addition to theory, the course also designs a complete big data practice course system that is compatible with theoretical courses, and provides an experimental manual on the cloud platform. The experimental case and the corresponding virtual machine environment enable students to build their own experimental environment or download the virtual machine environment according to the experimental manual while learning theoretical knowledge, and practice by themselves, understand the concepts and basic technologies related to big data, and experience each kind of algorithm and tools to experience the joy of big data analysis.

How are students assessed?
The course provides corresponding test questions and discussion questions for each unit, and at the same time provides final test questions for the mastery of test knowledge after all courses are taught. Take the unit test scores and final exam scores as the standard. The total score is 100 points, of which unit tests account for 30%, course experiments account for 20%, final exams account for 40%, and participation in discussions account for 10%.

Course tools and materials
Reference materials:

[1]. "Introduction to Big Data", edited by Mei Hong, Higher Education Press

[2]. "Introduction to Big Data", edited by Zhang Yaoxue, Machinery Industry Press
Several experiments such as tourism big data and data visualization are set up in the course, and an experiment manual is provided for everyone to practice on the cloud platform. You need to use the corresponding tools such as hadoop, java, python, etc., which are all free. The cloud platform used in the experiment will give each student 20 hours of free computer time each time.

Why do you teach the course this way?
This course is mainly in the form of classroom face-to-face, interspersed with small class discussions, flip classes, and experiments. Classroom face-to-face teaching is mainly taught by teachers, as the most common form of teaching, can be more comprehensive and systematic transfer of the main knowledge points to everyone. Small class discussion can promote communication and exchange between students, the classroom face-to-face process of problems to focus on solving, timely answer questions and puzzles, so that students have a clearer understanding of knowledge, while in the process of discussion can deepen their understanding of the corresponding knowledge points, in order to achieve the new effect of warm knowledge. Doing experiments can enhance the practical ability to operate, put the knowledge learned into use, not only on paper, general talk, the understanding of knowledge points more in place, more thorough.

Body of Knowledge coverage

KA	Sub-domain	Competencies Covered	Hours

AP	Visualization	Gain familiarity with the main strands of knowledge underpinning approaches to Analysis and Presentation Provide the range of skills and techniques that can be employed in addressing the challenges of analysis and presentation and creating efficient and effective interfaces	5
PDA	Data structures	Write clear and correct code in a programming language that includes primitive data types, references, variables, expressions, assignments, I/O, control structures, functions, and recursion. Use standard libraries for a given programming language. Select appropriate data structures for a given problem.	3
DM	Data Analysis	Data analysis method	7
BDS	Big Data Applications	Overview of Big Data Applications Smart medical big data Travel big data case	4
DAMG	Working with various types of data	Data acquisition and cleaning Data storage and management Data processing and analysis Resource management and scheduling	13