

ACM Data Science Task Force Course Example

*Big Data Computing Technology
University of Electronic Science and Technology of China, chengdu
Di Lin, Yu Tang*

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

Knowledge Area	Total Number of Contact Hours
Software support for Big Data applications	14
Data acquisition	2
Data cleaning	4
Working with various types of data	8
The basic concepts and models of data.	4

Where does the course fit in your undergraduate Data Science curriculum?

The nature of this course is a professional elective course, the first course is data structure and algorithms, database principles and applications, elective courses for software engineering professional training direction. This course is a basic course of data analysis for information technology related programs. In this course, students will establish a clear and comprehensive understanding of the value, meaning and basic principles of large-scaled data analysis, master the basic principles and methods of data mining, processing and modeling, study the computing architecture, computing model and development technologies in big data processing, acquire the methods of data analysis in health care, Finance, E-commerce, public management and the other application fields.

What is covered in the course?

Establish the fundamental concepts of big data and understand the key attributes and design principles. Understand the characters of various data analysis and master the analysis methods based on the programming language of Python and Java. Understand the design of distributed data computing and its implementation. Develop the preliminary capability of data analysis and its implementation through the lectures and a course project on the requirements of data analysis.

What is the format of the course?

This course consists of 4 hours of extracurricular labs, focusing on the use of the Hadoop platform and python development environment to complete a hands-on course design for big data analytics, requiring business requirements analysis, distributed data storage design, data cleaning and protocol method design, data analysis algorithm design, and data analysis reporting. The face-to-face course first introduces the basic concepts of big data, the basic elements and design principles of big data analysis, the pattern characteristics and design principles of various big data computing. Then the data acquisition method, data cleaning and analysis method, data processing and analysis technology are introduced. Including the source, method and step content of data acquisition, data cleaning and analysis method data cleaning,

data protocol, data analysis method, text reading and writing technology, data processing technology, data analysis technology, visualization technology, big data computing system architecture and so on. Finally, the paper introduces hadoop eco-cluster, MapReduce method, graph parallel computing framework, interactive computing framework, stream computing system, memory computing mode, Spark computing platform.

How are students assessed?

This course is based on a percentage system, and the assessment method of this course is the usual assessment (course experiment) (40%), the final assessment (60%). The final exam is an open-book exam.

Course tools and materials

Textbooks: "Big Data Analysis and Computing" by Tang Yu, Lin Di, Fan Aihua, Wu Weiwei, Tsinghua University Press, 1st edition, 2017.

Resources:

1. Data Analysis with Python, by McKinney, Machinery Industry Press, 1st edition, 2014
2. Algorithms in the Age of Big Data: Machine Learning, Artificial Intelligence and Its Typical Examples, edited by Fanping Liu, Electronic Industry Publishing Society, 1st edition, 2017.

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. 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
BDS	Software support for Big Data applications	Mass data processing platform Hadoop eco-cluster, MapReduce method, graph parallel computing framework, interactive computing framework, stream computing system, memory computing mode, Spark computing platform.	14
DG	Data acquisition, Data cleaning, Working with various types of data	<ol style="list-style-type: none"> 1. The basic concepts of big data, the basic elements and design principles of big data analysis, the pattern characteristics and design principles of various big data computing. 2. The source, method, and step content of the data acquisition. 3. Data cleaning and analysis methods Data cleaning, data protocol, data analysis methods. 4. Text reading and writing technology, data processing technology, data analysis technology, visualization 	18

		technology, big data computing system architecture.	
--	--	---	--

Additional topics

Other comments