ACM Data Science Task Force Course Example

DG, DM, Big Data Analytics
Harbin Institute of Technology, Harbin
Shengfei Shi, Donghua Yang, Zhaonian Zou

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

Knowledge Area	Total Number of Contact Hours
Continuing professional development	2
Data acquisition, Working with various types of data	8
Cluster analysis, Classification and regression	20
Big Data Analytics Method	20
Big data analysis performance optimization	8
Techniques for Big Data applications	2

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

The course of Big Data Analytics is designed to bring basic knowledge and methodology for big data analytics to students, and it is to make students familiar with recent development of big data analytics. The further influence of this course includes establishing a thinking towards big data analytics for students, and inspiring the capacity of dealing with big data problems for students. By providing professional materials and training, this course aims to improve students in their future career. The prerequisites for this course are data structure and algorithm, computer system, computer network, database system, software engineering, programming language design.

What is covered in the course?

The contents of this course are as follow: 1) the process of big data analytics, including data integration, data preprocessing, big data systems and evaluation of big data analytic results; 2) big data analytics in various fields, including text data, graph data and time series data; 3) multi-variate statistical analysis, including clustering analysis, discriminatory analysis, principle component analysis and factor analysis; 4) optimization of big data analytics, including optimization strategies based on algorithm design and system building. This course is to provide students with necessary technique and methodology on big data analytics, and make students capable of dealing with practical problems of big data in real life.

What is the format of the course?

The total number of hours in this course is 72 hours, of which 48 hours are taught in the large class and 24 hours are in the experiment. This is a course that combines theory and application. Classwork provides much of the content and expectations of the course. Curriculum related experiments are designed to allow students to better experience the practical application of the theories they have learned.

How are students assessed?

This course has a score of 100. It consists of:

1. Open big job (15%) ➢ Big job title: Big data analysis and performance evaluation based on Hadoop/Spark.

➢ Big job content: According to the current big data application scenarios, such as industrial, medical and
In the field of science, design a big data analysis application that requires coverage from requirements
Analyze the whole process of system design, implementation, and evaluation, and write a course report.

2. Experiment (15%)

➢ Decompose project development into experimental contents such as data preprocessing, statistical analysis, big data analysis and application, and big data analysis and optimization. A total of 4 experiments, 24 studies
Time.

3. Final exam (60%)

The final exam covers all teaching content, and the ratio is:

➢Basic concepts of big data analysis (15%); ➢Big data analysis process (10%); ➢Selected lectures on multivariate statistical analysis (30%); ➢Selected lectures on big data analysis in various fields (30%); ➢Big data analysis performance Optimization (15%)

4. Classroom lectures, homework (10%) and homework submitted to the corresponding course mailbox.

Course tools and materials

This course does not require any materials, the teacher will send the materials to everyone when teaching. In addition, the following two textbooks can be used as teaching reference for students to supplement reading:

1. Wang Hongzhi, Principles and Practice of Big Data Analysis, China Machine Press
2. Ren Xuesong, Yu Xiulin, Multivariate Statistical Analysis, China Statistics Press
3. EMC Education Services, Data Science and Big Data Analytics
4. Mahmoud Parsian, Data Algorithms

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
PR	Continuing professional development	Understand what big data analysis is	2

		• Definition, meaning, application scenarios, etc. • Understand the key technologies involved in big data analysis (*) • Data collection, data management, infrastructure, data understanding and extraction, statistical analysis, data mining, data visualization, etc. • Understand the difficulties of big data analysis (*) • Scalability, usability, combination with specific domain knowledge, result verification • Understand the cutting-edge results of big data analysis	
DG	Data acquisition, Working with various types of data	• Collection and storage of big data • Big data preprocessing (*) • Big data analysis and modeling (*) • Big data analysis method • Big data analysis results display and evaluation • Able to use Hadoop/Spark platform to program data integration and data preprocessing methods.	8
DM	Cluster analysis, Classification and regression	• Cluster analysis • Classification analysis • Association analysis • Big data dimensionality reduction • Able to use Hadoop/Spark platform to program cluster analysis, classification analysis, association analysis, etc.; use advanced programming language to program cluster analysis, classification analysis, association analysis and other algorithms, and use multivariate statistical tool SSPS to test.	20
BDS	Techniques for Big Data applications	• Social Network Service • Recommender System (Recommender System)	2

Additional topics
Big Data Analytics Method, Big data analysis performance optimization