ACM Data Science Task Force Course Example

Big Data Technology Principle and Citation

Xiamen University,Xiamen
Ziyu Lin
ziyulin@xmu.edu.cn

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

Knowledge Area	Total Number of Contact Hours
Big data computing architectures	4
Distributed data storage	8
Techniques for Big Data applications	8
Software support for Big Data applications	12

Where does the course fit in your undergraduate Data Science curriculum?
This course is positioned as an entry-level big data course, starting with object-oriented programming (such as Java), databases, and operating systems. The majority of the students who take this course are engineering students in the middle grade, with slightly more boys.

Is this course from or used in other curricula/majors?

What is covered in the course?
This course systematically discusses the basic concept of big data and the architecture of big data processing. Hadoop, Distributed File System HDFS, Distributed Database HBase, NoSQL Database, Cloud Database, Distributed Parallel Programming Model. MapReduce, Spark, Stream Computing, Graph Computing, Data Visualization, and Big Data are used in the Internet, biomedical, and logistics. In important chapters such as Hadoop, HDFS, HBase, MapReduce, Spark, and more, entry-level hands-on operations are scheduled to enable students to better learn and master key big data technologies.

What is the format of the course?
The total number of hours in this course is 32 hours. This course introduces big data-related knowledge. With the guiding ideology of "building knowledge system, clarifying basic principles, guiding primary practice and understanding related applications", the course systematically combs the big data knowledge system, so as to achieve "orderly organization, roughing, shallowness and gradual development", and aims to build a bridge and link to "big data knowledge space" for students. The course will systematically summarize the technology related to big data, introduce the basic principles of big data technology and the main applications of big data, help students to form a contoured understanding of big data knowledge system and its application areas, and lay the foundation and direction for students to "in-depth study" in the field of big data.

How are students assessed?
At the end of each module, students are required to take the unit test, which includes single-choice questions, multiple-choice questions, judgment questions, and fill-in questions. Ultimately, the total course score consists of a unit test (50%) and an exam (50%). The total score includes "unqualified", "qualified" and "excellent" three grades, the total score is less than 60 points, is "unqualified", greater than or equal to 60 points less than 85 points is "qualified", greater than or equal to 85 points is "excellent".

Course tools and materials
[1] Familiarize yourself with commonly used Linux and Hadoop operations.
[2] Familiar with the commonly used HDFS operation.
[3] Familiar with the commonly used HBase operation.
[4] NoSQL and relationship database operation comparison.
[5] MapReduce Primary Programming Practice.
[6] Principles and Applications of Big Data Technology (3rd Edition), Lin Ziyu. People's Posts and Telecommunications Press
[7] Introduction to Big Data (General Education Edition), Lin Ziyu. Higher Education Press
[8] Big data basic programming, experiment and case tutorial, Lin Ziyu, Tsinghua University Press
[9] Basics of Spark Programming (Scala Edition), Lin Ziyu, Lai Yongxuan, Tao Jiping. People's Posts and Telecommunications Press

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.

Body of Knowledge coverage

KA	Sub-domain	Competencies Covered	Hours
BDS	Big data computing architectures, Distributed data storage, Techniques for Big Data applications, Software support for Big Data applications	1. Big Data Overview. 2. Hadoop, the big data processing architecture. 3. Distributed file system HDFS. 4. Distributed database HBase. 5. NoSQL database. 6. Cloud database. 7. Mapreduce. 8. The Hadoop architecture is discussed again. 9. Data warehouse Hive. 10. Spark. 11. Stream calculation. 12. Flink. 13. Figure calculation. 14. Big data is used in different areas.	32