國立東華大學
教學計劃表　Syllabus

課程名稱(中文) Course Name in Chinese	大數據系統			學年/學期 Academic Year/Semester	106/2
課程名稱(英文) Course Name in English	Big Data Systems				
科目代碼 Course Code	CSIE59830	系級 Department & Year	碩士	開課單位 Course-Offering Department	資訊工程學系
修別 Type	選修 Elective	學分數/時間 Credit(s)/Hour(s)		3.0/3.0	
授課教師 Instructor	/吳秀陽				
先修課程 Prerequisite					

課程描述　Course Description

The term "big data" is now commonly used to mean that the growth of data in volume, velocity, variety and veracity are in such an unprecedented scale that traditional database management systems can no longer handle it properly. Take Walmart – the world's biggest retailer with over 20,000 stores in 28 countries, as an example, is in the process of building the world's biggest private cloud, to process 2.5 petabytes of data every hour. Facebook, the world's most popular social media network, needs to process data from more than 2 billion monthly active users worldwide. Every 60 seconds, 136,000 photos are uploaded, 510,000 comments are posted, and 293,000 status updates are posted. That amounts to 1000+ terabytes of data generated per day. Approximately 600 million times per second, particles collide within the Large Hadron Collider (LHC) at CERN. Merely recording these events takes up 500EB(1EB = 1024PB) of storage per day, let alone analyzing it. We therefore need new technologies (big data processing) and new tools (big data systems) for these jobs. This is an introductory course on big data concepts, processing, analytics and systems. You will learn the latest development in big data technologies and get hands on experience in using popular open source big data systems such as Hadoop, HBase, Spark, Hama, etc.

The objectives of this course can be summarized as follows.
1. Understand big data concepts, challenges and trends.
2. Learn the technological foundations of big data science & engineering.
3. Learn the principles and practices behind popular open source big data systems.
4. Get hands on experiences of using open source big data systems for solving big data problems.

This is a lecture-oriented course. The system part of the course will be executed through in-class example discussion, homework assignments and term project. Due to the time limit, the lectures will focus mostly on the technological innovation of each system rather than how to use them. With brief introduction to the basic operations of various big data systems, students are expected to learn to use them on their own.

課程目標　Course Objectives

本課程的目標在於讓修課同學：充分了解Google等國際頂尖公司的技術與應用發展脈絡，做好迎接巨量資料時代的準備，掌握知識經濟下，全球化，數位化，以及網路化時代下的科技人必備的競爭利器。

	系專業能力 Basic Learning Outcomes	課程目標與系專業能力相關性 Correlation between Course Objectives and Dept.'s Education Objectives
A	統合資工知識技術之能力Ability to integrate knowledge and technologies of computer science and information engineering.	●

B	設計技術理論驗證實驗之能力Ability to design and conduct science experiments and to validate hypotheses.	●
C	資訊軟硬體設計開發之能力Ability to design and develop computer software and hardware.	●
D	團隊專案開發之能力Ability to design and develop team projects.	○
E	批判性思考與創新研發之能力Ability of analytical thinking, creative research planning, and innovative development.	●

圖示說明Illustration ：● 高度相關 Highly correlated ○中度相關 Moderately correlated

授 課 進 度 表 Teaching Schedule & Content		
週次Week	內容 Subject/Topics	備註Remarks
1	Course description, syllabus, and introduction	
2	General purpose big data systems I 1. Distributed and cluster computing 2. Apache Hadoop	
3	General purpose big data systems II 1. MapReduce 2. MapReduce on Hadoop	
4	General purpose big data systems III 1. In-memory computation 2. Apache Spark	
5	Big data storage I 1. Distributed filesystems and big data storage 2. Google GFS	
6	Big data storage II 1. Apache HDFS 2. Google BigTable system	
7	Big structured data processing I 1. SQL or NoSQL 2. Apache HBase	
8	Big structured data processing II 1. Cassandra and MongoDB 2. Data Warehousing, Google BigQuery and Apache Hive	
9	期中考試週 Midterm No midterm. Independent study group topic proposal.	
10	Big graph processing I 1. The challanges of big graphs 2. Pregel family of systems	
11	Big graph processing II 1. GraphLab family of systems 2. Applications of big graph computation	
12	Big stream processing I 1. The challenges of distributed big stream processing 2. Apache Flink	
13	Big stream processing II 1. Apache Storm 2. Spark Streaming	
14	Other systems and trends** 1. Big data analytics 2. Google Dremel, Apache Drill and Apache Impala 3. Google Cloud Platform (GCP) vs Amazon Web Services (AWS) 4. Open Data 6. Beyond Hadoop	
15	Independent study presentation I	
16	Independent study presentation II	
17	No class on the Dragon Boat Festival	

18	期末考試週 Final Exam		

教　學　策　略　Teaching Strategies

☑ 課堂講授 Lecture　　　☐ 分組討論Group Discussion　　　☐ 參觀實習 Field Trip

☐ 其他Miscellaneous:

學期成績計算及多元評量方式 Grading & Assessments

配分項目 Items	配分比例 Percentage	多元評量方式 Assessments							
		測驗會考	實作觀察	口頭發表	專題研究	創作展演	卷宗評量	證照檢定	其他
平時成績 General Performance	15%			✓	✓				
期中考成績 Midterm Exam									
期末考成績 Final Exam	25%	✓							
作業成績 Homework and/or Assignments	35%		✓						
其他 Miscellaneous (Term project)	25%		✓			✓			

評量方式補充說明
Grading & Assessments Supplemental instructions

教科書與參考書目（書名、作者、書局、代理商、說明）
Textbook & Other References (Title, Author, Publisher, Agents, Remarks, etc.)

No required textbook.　The following references are useful:
1. Kai Hwang and Min Chen. Big Data Analytics for Cloud, IoT and Cognitive Computing. John Wiley & Sons Ltd., 2017.
2. Sherif Sakr. Big Data 2.0 Processing Systems: A Survey. Springer, 2016.
3. Tom White. Hadoop: The Definitive Guide, 4th Edition, O'reilly, 2015.
4. Jure Leskovec, Anand Rajaraman, Jeff Ullman. Mining of Massive Datasets. Cambridge University Press, 2010~2014.
5. Mohammed J. Zaki and Wagner Meira JR. Data Mining and Analysis – Fundamental Concepts and Algorithms. Cambridge University Press, 2014.
6. Donald Miner and Adam Shook. MapReduce Design Patterns, O'reilly, 2013.

課程教材網址（教師個人網址請列在本校內之網址）
Teaching Aids & Teacher's Website (Personal website can be listed here.)

http://web.csie.ndhu.edu.tw/showyang/BigDataSys2018s/index.html

其他補充說明（Supplemental instructions）