國立東華大學

課　綱　Course Outline
資訊工程學系資工組

中文課程名稱 Course Name in Chinese	巨量資料系統				
英文課程名稱 Course Name in English	Big Data Systems				
科目代碼 Course Code	CSIE59830		班　別 Degree	碩士班 Master's	
修別 Type	選修 Elective	學分數 Credit(s)	3.0	時　數 Hour(s)	3.0
先修課程 Prerequisite					

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。

系教育目標 Dept.'s Education Objectives

1	探究學科知識，善用專業技能 Explore academic knowledge, utilize professional skills.
2	訓練評析思考，創新解決問題 Exercise analytical thinking, enhance creative problem solving skills.
3	學習團隊分工，強化溝通表 Participate in teamwork, strengthen communication skills.

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

課程大綱
Course Outline

1.　　　Introduction：Data -> Knowledge -> Intelligence，Roadmap. What/Why/When of Big Data. Review of the opportunities and challenges for Big Data. Data Warehousing。
2.　　　Computation (I)：MapReduce。
3.　　　Computation (II)：MapReduce & Practice for Apache Hadoop。
4.　　　In-Memory Computation：Shark/Spark/Sparrow。
5.　　　Graph Computation：Introduction to the new graph-based systems and Google's Pregel. Practice and hand-on for Apache's new graph-based system (Hama)。
6.　　　Storage (I)：Distributed filesystems and Google's GFS。
7.　　　Storage (II)：Distributed storage and Google's BigTable system。
8.　　　Storage (III)：Practice and hand-on for Apache HDFS and Hbase。
9.　　　Data Warehousing。
10.　　　Analytics (I)：Google's Dremel and Apache's interactive query system (Drill)。
11.　　　Analytics (II)：Big Explorer and machine learning, algorithms and applications。
12.　　　Technology Trends：Knowledge Graph, Open Data, Beyond Hadoop。

資源需求評估（師資專長之聘任、儀器設備的配合‧‧‧等）
Resources Required (e.g. qualifications and expertise, instrument and equipment, etc.)

須搭配電腦教室與一般教室使用。

課程要求和教學方式之建議
Course Requirements and Suggested Teaching Methods

Backgrounds in programming, file system and database and operating systems are highly desirable。

其他
Miscellaneous

Textbook："The Datacenter as a Computer", Urs Holzle and Luiz Andre Barroso。
Reading requirements：
1.　　　"Hadoop: The Definitive Guide", 2nd Edition. Tom White。
2.　　　"Mining of Massive Data Sets", Anand Rajaraman, Jure Leskovec and Jeff Ullman. Available at http://infolab.stanford.edu/~ullman/mmds.html#latest 。