

Trường Đại học Bách Khoa -
ĐHQG Tp.HCM
Khoa: Khoa Khoa học và Kỹ
thuật Máy tính
Khoa/Bộ môn quản lý MH: Văn
phòng khoa

Tp.HCM, ngày tháng
năm

Đề cương môn học Sau đại học

DỮ LIỆU LỚN (BIG DATA)

Mã số MH: CO5153

Số tín chỉ:	Tc (LT.BT&TH.Tự Học): 3						TCHP:					
Số tiết	-Tổng:	75	LT:	30	BT:	0	TH:	0	ĐA:		BTL/TL:	45
Đánh giá:	Thuyết trình:			20%								
	Tiểu luận:			30%								
	Thi cuối kỳ:			50%								
- Môn tiên quyết:												
- Môn học trước:												
- Môn song hành:												
- CTĐT ngành (Mã ngành):	Khoa Học Máy Tính (8480101), Hệ Thống Thông Tin Quản Lý (8340405)											
- Ghi chú khác:												

1. Mục tiêu môn học:

Môn học là giúp cho học viên học, hiểu và giải quyết các vấn đề liên quan đến dữ liệu lớn. Mục tiêu chính của môn học: khái niệm tổng quan về dữ liệu lớn, dữ liệu và dữ liệu lớn, giải thuật và công nghệ tính toán về dữ liệu lớn.

Aims:

The main goal of this course is to help students learn, understand, and solve problems of big data. Mainly the course objectives are: conceptualization and summarization of big data, trivial data versus big data, big data algorithms and computing technologies.

2. Nội dung tóm tắt môn học:

Dữ liệu lớn và các ứng dụng liên quan từ góc nhìn kinh doanh và kỹ thuật. Các đặc trưng cơ bản về Dữ liệu lớn: dung lượng lớn, tốc độ sinh dữ liệu lớn, tính đa dạng và không đầy đủ của dữ liệu. Giới thiệu các kỹ thuật, công cụ giải các bài toán về Dữ liệu lớn.

Course outline:

Big data and its applications from the view of business and technology. The characteristics of Big data: volume, velocity, variety and veracity. Introduction to technologies and tools used in Big data applications.

3. Tài liệu học tập:

Sách giáo trình (Text books)

[1] Hrushikesh Mohanty, Prachet Bhuyan, Deepak Chenthati: Big Data - A Primer, Springer 2015.

[2] Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman: Mining of Massive Datasets, 2014.

[3] Parallel & distributed computing: https://computing.llnl.gov/tutorials/parallel_comp/#Whatis.

Sách tham khảo (References)

[4] Hadoop: <http://hadoop.apache.org>.

[5] Spark: <https://spark.apache.org/docs/latest/quick-start.html>.

4. Các hiểu biết, các kỹ năng cần đạt được sau khi học môn học:

STT	Chuẩn đầu ra môn học (CĐRMH)	Công cụ đánh giá CĐRMH	Đóng góp CDR Chương trình (CĐRCT)		
			Ứng dụng	Nghiên cứu	
CĐRMH.1	Hiểu về vai trò của Dữ liệu lớn và các ứng dụng có liên quan		a, c		1.1, 1.3
CĐRMH.2	Có khả năng sử dụng công cụ Hadoop, Spark để phát triển ứng dụng		b, c		1.3
CĐRMH.3	Có khả năng tự học và nghiên cứu các chủ đề mới có liên quan đến Dữ liệu lớn		d, e, g		1.4, 2.1, 2.2, 2.3, 2.4, 3.1, 4.1

Learning outcomes:

No.	Course learning outcomes (CLO)	CLO assessment	Matching with PLO		
			Coursework	Research	
L.O.1	Understand Big data and its applications		a, c		1.1, 1.3
L.O.2	Be able to use tools like Hadoop, Spark to develop applications		b, c		1.3
L.O.3	Study new topics in Big data		d, e, g		1.4, 2.1, 2.2, 2.3, 2.4, 3.1, 4.1

Bảng ánh xạ chuẩn đầu ra môn học và chuẩn đầu ra chương trình ứng dụng:

Chuẩn đầu ra môn học (CĐRMH)	Chuẩn đầu ra của chương trình (CĐRCT)										
	a	b	c	d	e	f	g	h	i	j	k
CĐRMH.1	✓		✓								
CĐRMH.2		✓	✓								
CĐRMH.3				✓	✓		✓				

Bảng ánh xạ chuẩn đầu ra môn học và chuẩn đầu ra chương trình nghiên cứu:

Chuẩn đầu ra môn học (CĐRMH)	Chuẩn đầu ra của chương trình (CĐRCT)										
	a	b	c	d	e	f	g	h	i	j	k
CĐRMH.1											
CĐRMH.2											
CĐRMH.3											

5. Hướng dẫn cách học - chi tiết cách đánh giá môn học:

Tham dự giờ giảng trên lớp + làm bài tập đầy đủ

Về kiểm tra giữa kỳ (tuần kiểm tra, thời lượng....): không có kiểm tra giữa kỳ

Về thực hiện báo cáo tiểu luận: nhóm 2-4 người, đánh giá: 30%

Thuyết trình trên lớp: nhóm 2-4 người, đánh giá: 20%

Cách tổ chức thi cuối kỳ: 50%, viết và trắc nghiệm

Ghi chú về điều kiện cầm thi, cách tổng kết điểm: sinh viên không được thiếu bất cứ cột điểm nào của môn học mới được tính điểm sau cùng của môn

Learning strategies & Assessment Scheme:

Join lectures + do exercises

No midterm exam

Project (2-4 persons): 30%

Seminar (2-4 persons): 20%

Final exam: 50%

Notes: students have to involve all activities

6. Nội dung chi tiết:

Tuần/ Buổi	Chủ đề (chương)	Nội dung	Chuẩn đầu ra môn học	Tài liệu
1	Giới thiệu dữ liệu lớn	1.1 Dữ liệu lớn là gì? 1.2 Ứng dụng của dữ liệu lớn	CĐRMH.1	[1]
2	Tính toán song song & phân tán	2.1 Giới thiệu tính toán song song & phân tán 2.2 Kỹ thuật tính toán song song & phân tán	CĐRMH.1, CĐRMH.2	[3]
3	Map/Reduce	3.1 Giới thiệu Map/Reduce 3.2 Hadoop và ứng dụng	CĐRMH.2	[4, Internet]
4	Hệ thống lưu trữ	4.1 Hệ thống file phân tán: HDFS, GFS, Lustre 4.2 Hệ thống lưu trữ	CĐRMH.1, CĐRMH.2	[Internet]
5	Spark & data analytics	5.1 Introduction to Spark 5.2 Applications	CĐRMH.2	[Internet]
6	Streaming & data stream	6.1 Mô hình streaming 6.2 Kỹ thuật streaming 6.3 Ứng dụng	CĐRMH.1, CĐRMH.3	[1, Internet]
7	Clustering & Classification	7.1 Clustering 7.2 Classification	CĐRMH.1, CĐRMH.3	[2, Internet]

Tuần/ Buổi	Chủ đề (chương)	Nội dung	Chuẩn đầu ra môn học	Tài liệu
8	Graph Computing	8.1 Kỹ thuật Graph computing 8.2 Ứng dụng	CĐRMH.1, CĐRMH.3	[2, Internet]
9	Trực quan hoá dữ liệu lớn	9.1 Giới thiệu kỹ thuật trực quan hoá 9.2 Ứng dụng	CĐRMH.1, CĐRMH.3	[Internet]
10	Ôn tập	Ôn tập		

7. Giảng viên tham gia giảng dạy:

CBGD
chính:

PGS.TS
Thoại
Nam

CBGD
tham
gia:

TS.
Nguyễn
Quang
Hùng

**XÁC NHẬN
CỦA HỘI
ĐỒNG XÂY
DỰNG
CHƯƠNG
TRÌNH ĐÀO
TẠO VÀ KHOA**

*Tp. Hồ Chí
Minh, ngày
..... tháng
..... năm*

.....
**GIẢNG
VIÊN
LẬP ĐỀ
CƯƠNG**

**PGS.TS
Thoại
Nam**