

DỮ LIỆU LỚN (BIG DATA)

Mã số MH: 055214

Số tín chỉ:		Tc (LT.BT&TH.Tự Học): 3				TCHP:			
Số tiết	-Tổng:	60	LT: 45	BT: 0	TH: 0	ĐA:		BTL/TL:	15
Đánh giá:		Kiểm tra giữa kỳ:			20%				
		Tiểu luận:			30%				
		Kiểm tra cuối kỳ:			50%				
- Môn tiên quyết:									
- Môn học trước:									
- Môn song hành:									
- CTĐT ngành (Mã ngành):		Bản Đồ, Viễn Thám Và Hệ Thống Thông Tin Địa Lý (8440214)							
- Ghi chú khác:									

1. Mục tiêu môn học:

Cung cấp cho học viên kiến thức cơ bản và ứng dụng của dữ liệu lớn (Big data); giới thiệu các công cụ Hadoop, Spark và các giải pháp kỹ thuật để giải quyết các vấn đề liên quan đến Big data.

Aims:

This course looks at the foundation of Big Data and its applications; the focus will be on tools like Hadoop, Spark as well as technologies used to solve Big Data applications.

2. Nội dung tóm tắt môn học:

Big data và các ứng dụng liên quan từ góc nhìn kinh doanh và kỹ thuật. Các đặc trưng cơ bản về Big data: 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 Big data.

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:

- [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.
- [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:

Học viên hiểu về vai trò của Big data và các ứng dụng có liên quan, có khả năng sử dụng công cụ Hadoop, Spark để phát triển ứng dụng, có khả năng tự học và nghiên cứu các chủ đề mới có liên quan đến Big data.

Learning outcomes:

At the end of the course, students will understand Big data and its applications, be able to use tools like Hadoop, Spark to develop applications and study new topics in Big data.

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 cấm 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	Tài liệu
1	Giới thiệu dữ liệu lớn	1.1 Dữ liệu lớn là gì? 1.2 Ứng dụng của dữ liệu lớn	[1]
2	Xử lý song song & phân tán	2.1 Giới thiệu tính toán song song & phân tán 2.2 Kỹ thuật tính toán song song & phân tán	[4]

Tuần /Buổi	Chủ đề (chương)	Nội dung	Tài liệu
3	Map/Reduce	3.1 Giới thiệu Map/Reduce 3.2 Hadoop và ứng dụng	[5, Internet]
4	Hệ thống lưu trữ	4.1 Hệ thống file phân tán: HDFS, GFS, Lustre 4.2 Hệ thống lưu trữ	[Internet]
5	Spark & phân tích dữ liệu	5.1 Giới thiệu công cụ Spark 5.2 Ứng dụng	[5, Ineternet]
6	Streaming & data stream	6.1 Mô hình streaming 6.2 Kỹ thuật streaming 6.3 Ứng dụng	[3, Internet]
7	Clustering & Classification	7.1 Clustering 7.2 Classification	[2, Internet}
8	Graph Computing	8.1 Kỹ thuật Graph computing 8.2 Ứng dụng	[2, Internet]
9	Trực quan hoá dữ liệu lớn	9.1 Giới thiệu kỹ thuật trực quan hoá 9.2 Ứng dụng	[Internet]
10	Ôn tập		

7. Giảng viên tham gia giảng dạy:

CBGD chính: PGS.TS Thoại Nam

CBGD tham gia: TS. Nguyễn Quang Hùng

BỘ MÔN QUẢN LÝ MÔN HỌC

Tp. Hồ Chí Minh, ngày tháng năm

GIẢNG VIÊN LẬP ĐỀ CƯƠNG