

CHƯƠNG TRÌNH ĐÀO TẠO TRÌNH ĐỘ ĐẠI HỌC
NGÀNH ĐÀO TẠO: KHOA HỌC DỮ LIỆU VÀ TRÍ TUỆ NHÂN TẠO

TH94013: XỬ LÝ DỮ LIỆU LỚN (BIG DATA)

I. Thông tin về học phần

- Học kỳ: 6
- Tín chỉ: **3 (Lý thuyết 2 - Thực hành 1 - Tự học: 9)**
- Giờ tín chỉ đối với các hoạt động học tập
 - + Học lý thuyết trên lớp: 30 tiết
 - + Làm bài tập trên lớp: 10 tiết
 - + Thuyết trình và thảo luận trên lớp: 5 tiết
- Tự học: 135 tiết (theo kế hoạch cá nhân hoặc hướng dẫn của giảng viên)
- Đơn vị phụ trách:
 - Bộ môn: Khoa học máy tính
 - Khoa: Công nghệ thông tin
- Học phần thuộc khối kiến thức:

Đại cương ☐		Cơ sở ngành ☒		Chuyên sâu 1 ☐		Chuyên sâu 2 ☐		Chuyên sâu 3 ☐	
Bắt buộc	Tự chọn	Bắt buộc	Tự chọn	Bắt buộc	Tự chọn	Bắt buộc	Tự chọn	Bắt buộc	Tự chọn
☐	☐	☐	☒	☐	☐	☐	☐	☐	☐

- Ngôn ngữ giảng dạy: tiếng Anh ☐ Tiếng Việt ☒

II. Chuẩn đầu ra của chương trình đào tạo và mục tiêu, kết quả học tập mong đợi của học phần

** Các chuẩn đầu ra và chỉ báo của chương trình đào tạo mà học phần đóng góp:*

Chuẩn đầu ra Sau khi hoàn tất chương trình, sinh viên có thể:	Chỉ báo đánh giá việc thực hiện được chuẩn đầu ra
Kiến thức chung	
CĐR 2	2.1. Phân tích được các vấn đề cơ bản về công nghệ liên quan đến xây dựng hệ thống thông tin.
Kiến thức chuyên môn	
CĐR 3	3.1. Xây dựng được hệ thống xử lý và phân tích dữ liệu.
Kỹ năng chung	
CĐR 5	5.1. Vận dụng kỹ năng tư duy logic, kỹ năng phản biện để giải quyết hoặc đánh giá các vấn đề chuyên môn.
Kỹ năng chuyên môn	
CĐR 7	7.1. Phát triển hệ thống khai phá dữ liệu để giải quyết các vấn đề trong nhiều lĩnh vực.
Tự chủ và trách nhiệm	
CĐR 8	8.1. Thể hiện thói quen tự học, tự nghiên cứu, nâng cao trình độ chuyên môn.
CĐR 9	9.2. Có trách nhiệm đối với cá nhân và tập thể.

*** Mục tiêu:**

- Về kiến thức: Học phần cung cấp cho sinh viên tổng quan kiến thức, các khái niệm, đặc trưng cũng như những thách thức và công nghệ liên quan đến dữ liệu lớn. Đồng thời môn học giới thiệu một số phương pháp và công cụ phổ biến để khai thác và phân tích dữ liệu lớn (Hadoop, MapReduce và Spark).
- Về kỹ năng: Học phần rèn luyện cho sinh viên kỹ năng sử dụng một số công cụ trong việc triển khai và lập trình trên dữ liệu lớn.
- Năng lực tự chủ và trách nhiệm: Học phần rèn luyện cho sinh viên ý thức về trách nhiệm cá nhân đối với việc học tập, tuân thủ các yêu cầu, các quy định chung; Thái độ hợp tác trong nhóm; làm việc theo nhóm, nghiêm túc, chủ động tìm hiểu tích lũy kiến thức và nâng cao trình độ bản thân.

*** Kết quả học tập mong đợi của học phần:**

Học phần đóng góp cho Chuẩn đầu ra sau đây của CTĐT theo mức độ sau:

I – Giới thiệu (Introduction); P – Thực hiện (Practice); R – củng cố (Reinforce); M – Đạt được (Master)

Mã HP	Tên HP	Mức độ đóng góp của học phần cho chỉ báo CĐR của CTĐT					
		2.1	3.1	5.1	7.1	8.1	9.2
TH94013	Xử lý dữ liệu lớn	R	M	P	M	P	P

Ký hiệu	KQHTMD của học phần Hoàn thành học phần này, sinh viên thực hiện được	CĐR của CTĐT
Kiến thức chuyên môn		
K1	Trình bày công cụ và một số kỹ thuật dùng trong phân tích dữ liệu lớn.	2.1. Phân tích được các vấn đề cơ bản về công nghệ liên quan đến xây dựng hệ thống thông tin.
K2	Trình bày ứng dụng của phân tích dữ liệu lớn.	3.1. Xây dựng được hệ thống xử lý và phân tích dữ liệu.
Kỹ năng chung		
K3	Lựa chọn và phân tích mô hình ứng dụng cho bài toán xử lý dữ liệu cụ thể	5.1. Vận dụng kỹ năng tư duy logic, kỹ năng phản biện để giải quyết hoặc đánh giá các vấn đề chuyên môn.
Kỹ năng chuyên môn		
K4	Phân tích kết quả của phân tích dữ liệu lớn và đưa ra kết luận tối ưu cho từng trường hợp cụ thể.	7.1. Phát triển hệ thống khai phá dữ liệu để giải quyết các vấn đề trong nhiều lĩnh vực.
Tự chủ và trách nhiệm		
K5	Sử dụng thành thạo ngôn ngữ Python để phân tích xử lý một số bài toán về phân tích dữ liệu lớn.	8.1. Thể hiện thói quen tự học, tự nghiên cứu, nâng cao trình độ chuyên môn.
K6	Hình thành khả năng làm việc độc lập, cần cù, sáng tạo, chính xác, tinh thần trách nhiệm, hợp tác cao trong lập trình giải quyết các bài toán xử lý dữ liệu.	9.2. Có trách nhiệm đối với cá nhân và tập thể.

III. Nội dung tóm tắt của học phần

TH94013. Xử lý dữ liệu lớn(3TC: 2 –1 – 9). Mô tả vắn tắt nội dung: Học phần gồm 5 chương với nội dung: Tổng quan về dữ liệu lớn; Hadoop; MapReduce (MR); Spark; Phân tích dữ liệu lớn và Ứng dụng dữ liệu lớn.

IV. Phương pháp giảng dạy và học tập

1. Phương pháp giảng dạy

- Thuyết giảng và giảng dạy thông qua thực hành trực tiếp trên giảng đường hoặc trực tuyến thông qua Ms Teams.

2. Phương pháp học tập

- Sinh viên nghe giảng, kết hợp với tự học và trao đổi với bạn học và thầy cô trên giảng đường hoặc trực tuyến thông qua Ms Teams.

V. Nhiệm vụ của sinh viên

- Chuyên cần: Theo quy định chung của Học viện.

- Chuẩn bị cho bài giảng: Tất cả sinh viên tham dự học phần này phải đọc sách tham khảo trước khi đến lớp.

- Bài tập: Tất cả sinh viên tham dự học phần này phải hoàn thành ít nhất 50% lượng bài tập được giao.

- Thi giữa kì: Tất cả các sinh viên tham dự học phần này đều phải tham gia thi giữa kỳ.

- Thi cuối kì: Tất cả các sinh viên tham dự học phần này đều phải tham gia thi cuối kỳ.

VI. Đánh giá và cho điểm

1. Thang điểm: 10

2. Điểm trung bình của học phần là tổng điểm của các rubric nhân với trọng số tương ứng của từng rubric:

- **Đánh giá quá trình:**

+ Chuyên cần: 10 %

+ Bài tập giữa kỳ: 30%

- **Đánh giá cuối kì:**

+ Thi cuối kì: 60%

Bảng 1: Kế hoạch đánh giá và trọng số

Hoạt động đánh giá	KQHTMD được đánh giá	Trọng số (%)	Thời gian/Tuần học
Đánh giá quá trình			
Tham dự lớp	K1-K6	10	1-15
Bài tập giữa kỳ	K1-K4	30	13
Đánh giá cuối kì			
Thi cuối kỳ	K2-K4	60	Theo lịch

Bảng 2. Chỉ báo thực hiện các kết quả học tập mong đợi của học phần

Nội dung kiểm tra	Chỉ báo thực hiện của học phần được đánh giá qua câu hỏi	KQHTMD của môn học được đánh giá qua câu hỏi
Công cụ, kỹ thuật, chiến lược của dữ liệu lớn Hadoop, MapReduce, Spark	Chỉ báo 1: Trình bày công cụ trong phân tích dữ liệu	K1
	Chỉ báo 2: Trình bày một số kỹ thuật dùng trong phân tích dữ liệu lớn	K2
Một số ứng dụng của dữ liệu lớn.	Chỉ báo 3: Trình bày được ứng dụng của phân tích dữ liệu lớn	K3
	Chỉ báo 4: Trình bày được nhược điểm của phân tích dữ liệu lớn	K4

Phân tích kết quả và đưa ra kết luận trong từng trường hợp cụ thể	Chi báo 5: Phân tích kết quả của yêu cầu bài toán phân tích dữ liệu lớn	K4
	Chi báo 6: Đưa ra kết luận trong từng trường hợp cụ thể khi có kết quả phân tích dữ liệu lớn	K5

4. Các yêu cầu, quy định đối với học phần

Nộp bài tập chậm: Tất cả các trường hợp nộp bài tập chậm đều không được chấp nhận.

Tham dự các bài thi: Không tham gia bài thi giữa kỳ sẽ không được thi cuối kỳ. Trong trường hợp có lý do chính đáng sẽ được giảng viên bố trí cho kiểm tra bù (**trừ bài thi kết thúc học phần**)

Yêu cầu về đạo đức: Sinh viên có thái độ học tập nghiêm túc.

VII. Giáo trình/ tài liệu tham khảo

*** Sách giáo trình/Bài giảng:**

- EMC Education Services, ed. *Data science and big data analytics: discovering, analyzing, visualizing and presenting data*. John Wiley & Sons, 2015

- Peter Buhlmann, Sara van de Geer (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. NXB Springer.

*** Tài liệu tham khảo khác:**

- O'Reilly Radar Team(2012). *Big data now*, NXB O'Reilly.

VIII. Nội dung chi tiết của học phần

Tuần	Nội dung	KQHTMD của học phần
1 1	Chương 1: Tổng quan về dữ liệu lớn	
	A/ Các nội dung chính trên lớp: (3 tiết) Nội dung GD lý thuyết: (3 tiết) 1.1 Dữ liệu lớn là gì 1.2 Các nét đặc trưng 1.3 Nguồn hình thành	K1
	Ứng dụng của dữ liệu lớn	
	B/ Các nội dung cần tự học ở nhà: (9 tiết) Sinh viên làm các bài tập tương ứng các mục đã nghe giảng trên lớp	K1
	Chương 2: Hadoop	
2-4	A/ Tóm tắt các nội dung chính trên lớp: (14 tiết) Nội dung GD lý thuyết: (9 tiết) 2.1 Giới thiệu về mô hình GFS (Google File System) 2.2 Lịch sử Hadoop 2.3 Giải pháp Hadoop cho việc quản lý và khai thác Big data 2.4 Hệ thống file lưu trữ và quản lý của Hadoop: HDFS (Hadoop Distributed FileSystem) 2.6 Hadoop I/O 2.7 Apache Hadoop	K2
	Nội dung giảng dạy thực hành: (5 tiết) Hadoop: Cách cài đặt và hoạt động: +) Standalone +) Pseudo-Distributed +) Fully-Distributed	

Tuần	Nội dung	KQHTMD của học phần
	+) Cách vận hành	
	B/ Các nội dung cần tự học ở nhà: (42 tiết) Sinh viên làm các bài tập tương ứng các mục đã nghe giảng trên lớp	K1, K2
	Chương 3: MapReduce (MR)	
5-6	A/ Tóm tắt các nội dung chính trên lớp: (14 tiết) Nội dung GD lý thuyết: (9 tiết) 3.1 Giới thiệu MapReduce 3.2 Lập trình MapReduce Nội dung giảng dạy thực hành: (5 tiết) Lưu trữ và phân tích dữ liệu cơ bản với mô hình Hadoop-MapReduce sử dụng Python trên các tập dữ liệu mẫu có sẵn (gen sinh học DNA, dữ liệu văn bản (text) hoặc dữ liệu chứng khoán)	K2, K4, K5
	B/ Các nội dung cần tự học ở nhà: (42 tiết) Sinh viên làm các bài tập tương ứng các mục đã nghe giảng trên lớp	K1-K5
8-9	Chương 4: Spark	
	A/ Tóm tắt các nội dung chính trên lớp: (14 tiết) Nội dung GD lý thuyết: (9 tiết) 4.1 Giới thiệu về Apache Spark 4.2 Các tiềm năng và thử thách của Spark trong lĩnh vực “Khoa học dữ liệu (Data Science)” 4.3 Tìm hiểu về RDDs Nội dung giảng dạy thực hành: (5 tiết) Cài đặt và cách vận hành	K2, K4, K5
	B/ Các nội dung cần tự học ở nhà: (42 tiết) Sinh viên làm các bài tập tương ứng các mục đã nghe giảng trên lớp	

IX. Yêu cầu của giảng viên đối với học phần

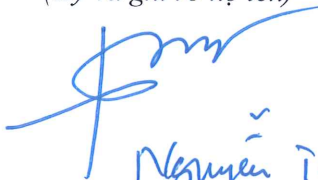
- Phòng học: đủ chỗ ngồi, đủ ánh sáng.
- Phương tiện phục vụ giảng dạy: có máy chiếu, micro.
- Các phương tiện khác: bảng viết phấn, phấn, khăn lau bảng.
- Elearning: phần mềm dạy và học trực tuyến MS Teams, máy tính, hệ thống máy chủ và hạ tầng kết nối mạng Internet với băng thông đáp ứng nhu cầu người dùng, không để xảy ra nghẽn mạng hay quá tải. Phòng học trực tuyến đầy đủ ánh sáng, cách âm tốt, thông thoáng, ngăn nắp, gọn gàng, sạch sẽ.

TRƯỞNG BỘ MÔN
(Ký và ghi rõ họ tên)


Trần Thị Thu Hương
KT. TRƯỞNG KHOA
(Ký và ghi rõ họ tên)


Ngô Công Chấn

Hà Nội, ngày 17 tháng 6 năm 2024
GIẢNG VIÊN BIÊN SOẠN
(Ký và ghi rõ họ tên)


Nguyễn Trung Kiên
KT. GIÁM ĐỐC
(Ký và ghi rõ họ tên)


PHÓ GIÁM ĐỐC
Phạm Văn Cường

PHỤ LỤC

THÔNG TIN VỀ ĐỘI NGŨ GIẢNG VIÊN GIẢNG DẠY HỌC PHẦN

Giảng viên phụ trách học phần

Họ và tên: Nguyễn Trọng Kương	Học hàm, học vị: Tiến sĩ
Địa chỉ cơ quan: Khoa CNTT, Học viện Nông nghiệp Việt Nam	Điện thoại liên hệ:
Email: ntkuong@vnua.edu.vn	Trang web: fita.vnua.edu.vn/ntkuong
Cách liên lạc với giảng viên: (nêu rõ hình thức liên lạc giữa sinh viên với giảng viên/trợ giảng)	

Giảng viên giảng dạy học phần

Họ và tên:	Học hàm, học vị:
Địa chỉ cơ quan:	Điện thoại liên hệ:
Email:	Trang web: (Đưa tên website của Khoa; website cá nhân – nếu có)
Cách liên lạc với giảng viên: (nêu rõ hình thức liên lạc giữa sinh viên với giảng viên/trợ giảng)	

Giảng viên giảng dạy học phần

Họ và tên:	Học hàm, học vị:
Địa chỉ cơ quan:	Điện thoại liên hệ:
Email:	Trang web: (Đưa tên website của Khoa; website cá nhân – nếu có)
Cách liên lạc với giảng viên: (nêu rõ hình thức liên lạc giữa sinh viên với giảng viên/trợ giảng)	

Giảng viên hỗ trợ /trợ giảng (nếu có)

Họ và tên:	Học hàm, học vị:
Địa chỉ cơ quan:	Điện thoại liên hệ:
Email:	Trang web: (Đưa tên website của Khoa; website cá nhân – nếu có)
Cách liên lạc với giảng viên: (nêu rõ hình thức liên lạc giữa sinh viên với giảng viên/trợ giảng)	

X. CÁC LẦN CẢI TIẾN (ĐỀ CƯƠNG ĐƯỢC CẢI TIẾN HÀNG NĂM THEO QUY ĐỊNH CỦA HỌC VIỆN)

- Lần 1: 07/2020
 - Xây dựng đề cương chi tiết mới để mở ngành
- Lần 2: 07/2021
 - Rà soát và cập nhật nội dung kiến thức học phần, rà soát tài liệu tham khảo
- Lần 3: 07/2022

- Rà soát và cập nhật tài liệu tham khảo
- Lần 4: 07/2023
 - Rà soát và cập nhật tài liệu tham khảo
- Lần 5: 07/2024
 - Cập nhật CDR của CTĐT
 - Cập nhật đóng góp bảng IPRM
 - Thay đổi mã môn học
 - Rà soát và cập nhật nội dung học phần và tài liệu tham khảo