

Đề Cương Chi Tiết Học Phần

1. Tên môn học: Phân tích dữ liệu lớn Mã học phần: BDAN333977

2. Tên Tiếng Anh: Big Data Analysis

3. Số tín chỉ: 3 (2+1) tín chỉ (2/1/6) (2 tín chỉ lý thuyết, 1 tín chỉ thực hành/thí nghiệm)

Phân bố thời gian: 15 tuần (2 tiết lý thuyết + 1*2 tiết thực hành + 6 tiết tự học/ tuần)

4. Các giảng viên phụ trách học phần:

1/ GV phụ trách chính: ThS. Hoàng Long

2/ Danh sách giảng viên cùng GD: TS. Lê Vĩnh Thịnh

5. Điều kiện tham gia học tập học phần

Môn học tiên quyết: Cơ sở dữ liệu (DBSY230184)

Môn học trước: Nhập môn môn dữ liệu lớn (BDES333877)

6. Mô tả học phần (Course Description)

Học phần cung cấp cho sinh viên kiến thức về kiến trúc của các hệ thống và các công cụ phục vụ cho hoạt động phân tích dữ liệu lớn. Với mỗi công cụ, môn học giới thiệu các kiến thức cơ bản và nâng cao cũng như phương thức tối ưu hóa hiệu suất hệ thống sử dụng công cụ này. Cùng với các bài tập lập trình, môn học hướng đến mục tiêu giúp người học có thể hình thành ý tưởng, thiết kế và hiện thực hóa hoạt động phân tích dữ liệu trong các hệ thống dữ liệu lớn.

7. Mục tiêu học phần (Course Goals)

Mục tiêu (Goals)	Mô tả (Goal description) (Học phần này trang bị cho sinh viên:)	Chuẩn đầu ra CTĐT	Trình độ năng lực
G1	Lập trình với Apache Pig, Hive, và Spark SQL.	ELO3(1.3)	3
G2	Phân tích, sử dụng hợp lý Apache Pig, Hive, và Spark SQL.	ELO4(2.1)	4
G3	Kỹ năng làm việc nhóm, giao tiếp và khả năng sử dụng các tài liệu kỹ thuật bằng tiếng Anh.	ELO9(3.1)	4
		ELO10(3.2)	5
		ELO11(3.3)	4
G4	Hình thành ý tưởng, thiết kế và hiện thực một phần của hệ thống dữ liệu lớn.	ELO12(4.1)	5
		ELO16(4.5)	5

8. Chuẩn đầu ra của học phần

Mục tiêu	Chuẩn đầu ra HP	Mô tả (Sau khi học xong môn học này, người học có thể:)	Chuẩn đầu ra CTĐT	Trình độ năng lực
----------	-----------------	--	-------------------	-------------------

G1	G 1.1	Lập trình với Apache Pig, Hive, và Spark SQL cho hệ thống dữ liệu lớn.	ELO3(1.3)	3
G2	G2.1	Phân tích và sử dụng hợp lý Apache Hive, Pig, và Spark SQL để giải quyết bài toán trong lĩnh vực dữ liệu lớn.	ELO4(2.1)	4
G3	G3.1	Làm việc theo nhóm. Trình bày và giao tiếp hiệu quả, trực tiếp cũng như bằng văn bản.	ELO9(3.1) ELO10(3.2)	4 5
	G3.2	Sử dụng hiệu quả các tài liệu kỹ thuật tiếng Anh phục vụ cho lĩnh vực.	ELO11(3.3)	4
G4	G4.1	Hình thành ý tưởng, thiết kế, và hiện thực hóa một phần hệ thống xử lý dữ liệu lớn sử dụng Pig, Hive, Spark SQL	ELO12(4.1) ELO16(4.5)	5 5

9. Đạo đức khoa học:

Các bài tập ở nhà và đồ án phải được thực hiện từ chính bản thân sinh viên. Nếu bị phát hiện có sao chép thì xử lý các sinh viên có liên quan bằng hình thức đánh giá **0** (không) điểm quá trình và cuối kỳ.

10. Nội dung chi tiết học phần:

Tuần	Nội dung	Chuẩn đầu ra học phần	Trình độ năng lực	Phương pháp dạy học	Phương pháp đánh giá
1-2	Chương 1: Kiến trúc hệ thống dữ liệu lớn (4/2/12)				
	A/ Các nội dung và PPGD chính trên lớp: (6) Nội Dung (ND) GD trên lớp 1.1. Dữ liệu lớn & hệ sinh thái dữ liệu lớn 1.2. Mô hình kiến trúc hệ thống dữ liệu lớn 1.3. Phân tích, so sánh kiến trúc một số hệ thống dữ liệu lớn điển hình	G2.1 G3.1 G3.2 G4.1	4 5 5 5	GV thuyết giảng, SV thuyết trình, Thảo luận	Bài thuyết trình, Project
	B/ Các nội dung cần tự học ở nhà: (12) + Tìm hiểu, phân tích kiến trúc hệ thống dữ liệu lớn + Thực hiện Project	G2.1 G3.1 G3.2 G4.1	4 5 4 5	Chuẩn bị bài thuyết trình, Thực hiện Project	
3-6	Chương 2: Apache Pig (8/4/24)				

	A/ Các nội dung và PPGD chính trên lớp: (12) Nội Dung (ND) GD trên lớp 2.1 Các kiến thức nền tảng về công cụ Pig. 2.2 Cơ bản về ngôn ngữ Pig Latin. 2.3 Ngôn ngữ Pig Latin nâng cao. 2.4 Tối ưu hóa hiệu suất hệ thống sử dụng Pig.	G1.1 G2.1 G3.1 G3.2 G4.1	3 4 5 4 5	GV thuyết giảng, SV thuyết trình, Thảo luận	Kiểm tra trên lớp, Bài tập về nhà, Project
	B/ Các nội dung cần tự học ở nhà: (24) + Bài tập lập trình Pig + Thực hiện Project	G1.1 G2.1 G3.1 G3.2 G4.1	3 4 5 4 5	Bài tập về nhà, Project	
7-10	Chương 3: Apache Hive (8/4/24)				
	A/ Các nội dung và PPGD chính trên lớp: (12) Nội dung GD lý thuyết: 3.1 Các kiến thức nền tảng về công cụ Hive 3.2 Cơ bản về ngôn ngữ HiveQL 3.3 Ngôn ngữ HiveQL nâng cao 3.4 Tối ưu hóa hiệu suất hệ thống sử dụng Hive	G1.1 G2.1 G3.1 G3.2 G4.1	3 4 5 4 5	GV thuyết giảng, SV thuyết trình, Thảo luận	Kiểm tra trên lớp, Bài tập về nhà, Project
	B/ Các nội dung cần tự học ở nhà: (24) + Bài tập lập trình Hive + Thực hiện Project	G1.1 G2.1 G3.1 G3.2 G4.1	3 4 5 4 5	Bài tập về nhà, Project	
11-14	Chương 4: Spark SQL (8/4/24)				
	A/ Các nội dung và PPGD chính trên lớp: (12) Nội dung GD lý thuyết: 4.1 Cơ bản về Spark SQL 4.2 Spark SQL nâng cao 4.3 Xây dựng ứng dụng sử dụng Spark SQL 4.4 Tối ưu hóa hiệu suất hệ thống sử dụng Spark SQL	G1.1 G2.1 G3.1 G3.2 G4.1	3 4 5 4 5	GV thuyết giảng, SV thuyết trình, Thảo luận	Kiểm tra trên lớp, Bài tập về nhà, Project
	B/ Các nội dung cần tự học ở nhà: (24) + Bài tập lập trình Spark SQL + Thực hiện Project	G1.1 G2.1 G3.1 G3.2 G4.1	3 4 5 4 5	Bài tập về nhà Project	
15	Báo cáo đồ án môn học (2/1/6)				

	A/ Các nội dung và PPGD chính trên lớp: (3) Nội dung GD lý thuyết: + Báo cáo Project	G1.1	3	Project	Project
		G2.1	4		
		G3.1	5		
		G3.2	4		
		G4.1	5		
	B/ Các nội dung cần tự học ở nhà: (6) + Chuẩn bị báo cáo Project	G1.1	3		
		G2.1	4		
		G3.1	5		
		G3.2	4		
		G4.1	5		

11. Đánh giá sinh viên:

- Thang điểm: 10

- Kế hoạch kiểm tra như sau:

Hình thức KT	Nội dung	Thời điểm	Chuẩn đầu ra đánh giá	Trình độ năng lực	Phương pháp đánh giá	Công cụ đánh giá	Tỉ lệ (%)
Bài tập lý thuyết							40
BT# 1	Thuyết trình về hệ thống dữ liệu lớn	Tuần 1-2	G2.1 G3.1 G3.2	2 5 4	Bài thuyết trình, Báo cáo	Rubric	10
BT# 2	Kiểm tra lý thuyết chương 2, 3, 4	Tuần 3 - 14	G1.1 G3.2	2 4	Kiểm tra trên lớp	Câu hỏi kiểm tra	30
Bài tập lớn							30
BL# 1	Bài tập lập trình chương 2, 3, 4	3 - 14	G1.1 G2.1 G4.1	3 3 5	Bài tập về nhà	Bài tập	30
Project							30
PR#1	Các nhóm sinh viên được yêu cầu đưa ra ý tưởng, thiết kế, hiện thực hóa một phần hệ thống phân tích dữ liệu lớn sử dụng các công cụ Pig, Hive và Spark SQL (Nhóm SV lựa chọn đề tài rồi thông báo cho GV, GV gợi ý, hướng dẫn SV cách làm). Bài tiểu luận được các nhóm SV báo cáo vào tuần cuối cùng.	Tuần 1-15	G1.1 G2.1 G3.1 G3.2 G4.1	3 4 5 4 5	Thực hiện project	Rubric	30

CDR	Hình thức kiểm tra
------------	---------------------------

