

**BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC SAO ĐỎ**

**ĐỀ CƯƠNG CHI TIẾT HỌC PHẦN
DỮ LIỆU LỚN - BIG DATA**

Số tín chỉ: 03

Trình độ đào tạo: Đại học

Ngành đào tạo: Công nghệ thông tin

Năm 2022

ĐỀ CƯƠNG CHI TIẾT HỌC PHẦN

Trình độ đào tạo: Đại học

Ngành đào tạo: Công nghệ thông tin

1. Tên học phần: Dữ liệu lớn - Big data

2. Mã học phần: CNTT 209

3. Số tín chỉ: 3 (2, 1)

4. Trình độ cho sinh viên: Năm thứ tư

5. Phân bổ thời gian

- Lên lớp: 30 tiết lý thuyết, 30 tiết thực hành.

- Tự học: 90 giờ.

6. Điều kiện tiên quyết: Không.

7. Giảng viên

STT	Học hàm, học vị, họ tên	Số điện thoại	Email
1	ThS. Phạm Thị Hương	0972.306.806	PTHuong@saodo.edu.vn
2	ThS. Nguyễn Thị Ánh Tuyết	0972.384.332	NTATuyet@saodo.edu.vn

8. Mô tả nội dung của học phần

Học phần Dữ liệu lớn - Big data giới thiệu tổng quan về khái niệm, đặc trưng cũng như những thách thức của Big data: Khả năng phân tích, dự đoán nhằm trích xuất một giá trị lớn hơn từ dữ liệu. Giới thiệu một số phương pháp và công cụ phổ biến để khai thác và quản lý Big data: Hadoop, MapReduce và Spark.

9. Mục tiêu và chuẩn đầu ra học phần

9.1. Mục tiêu

Mục tiêu học phần thỏa mãn mục tiêu của chương trình đào tạo:

Mục tiêu	Mô tả	Mức độ theo thang đo Bloom	Phân bổ mục tiêu học phần trong CTĐT
MT1	Kiến thức		
MT1.1	Giải thích phương pháp phân tích, xử lý một vấn đề cụ thể liên quan đến Big data.	2	[1.2.1.2b]
MT1.2	Minh họa cách sử dụng công cụ Hadoop HDFS để lưu trữ, mô hình MapReduce và Spark để phân tích dữ liệu lớn.	3	[1.2.1.2b]
MT1.3	Xác định được cách triển khai ứng dụng Big data trong thực tế.	3	[1.2.1.2b]

Mục tiêu	Mô tả	Mức độ theo thang đo Bloom	Phân bổ mục tiêu học phần trong CTĐT
MT2	Kỹ năng		
MT2.1	Áp dụng công cụ Hbase, Hadoop HDFS để lưu trữ, mô hình MapReduce và Spark để phân tích dữ liệu lớn.	3	[1.2.2.1]
MT2.2	Phân tích, tổng hợp, đánh giá các công cụ phân tích trong lĩnh vực xử lý dữ liệu lớn.	4	[1.2.2.1]
MT3	Mức tự chủ và trách nhiệm		
MT3.1	Nghiêm túc, tự giác, tích cực, khoa học, độc lập, cẩn thận và tuân thủ trong công việc.	3	[1.2.3.1]
MT3.2	Giải quyết được việc xử lý dữ liệu trong thực tế bằng Big data.	4	[1.2.3.2]

9.2. Chuẩn đầu ra

Sự phù hợp của chuẩn đầu ra học phần với chuẩn đầu ra của chương trình đào tạo:

CĐR học phần	Mô tả	Thang đo Bloom	Phân bổ CĐR học phần trong CTĐT
CĐR1	Kiến thức		
CĐR1.1	Giải thích được khái niệm, các đặc trưng cơ bản liên quan đến Big data.	2	[2.1.5]
CĐR1.2	Minh họa được các bước lưu trữ dữ liệu lớn bằng công cụ Hbase, Hadoop HDFS.	3	[2.1.5]
CĐR1.3	Phác họa được các bước phân tích dữ liệu lớn bằng mô hình MapReduce và Spark.	3	[2.1.5]
CĐR2	Kỹ năng		
CĐR2.1	Áp dụng công cụ, mô hình để lưu trữ, phân tích và triển khai được dữ liệu lớn.	3	[2.2.3]
CĐR2.2	So sánh được Hbase, Hadoop HDFS, MapReduce và Spark để đáp ứng các tình huống thực tế trong lĩnh vực xử lý dữ liệu lớn.	4	[2.2.3]
CĐR3	Mức tự chủ và trách nhiệm		
CĐR3.1	Nghiêm túc, tự giác, tích cực, khoa học, độc lập, cẩn thận, tuân thủ trong lập trình và thực tế công việc.	3	[2.3.1]

CDR học phần	Mô tả	Thang đo Bloom	Phân bố CDR học phần trong CTĐT
CDR3.2	Định hướng, hướng dẫn và đưa ra kết luận liên quan đến công việc phân tích và xử lý dữ liệu lớn.	4	[2.3.2]

10. Ma trận liên kết nội dung với chuẩn đầu ra học phần

Chương	Nội dung học phần	Chuẩn đầu ra của học phần						
		CDR1			CDR2		CDR3	
		CDR 1.1	CDR 1.2	CDR 1.3	CDR 2.1	CDR 2.2	CDR 3.1	CDR 3.2
1	Chương 1. Tổng quan về Big data 1.1. Khái niệm Big data 1.2. Các kiểu Big data 1.3. Các đặc trưng của Big data	2			2		3	
2	Chương 2. Sử dụng Hbase 2.1. Giới thiệu 2.2. Các tính năng 2.3. Mô hình 2.4. Kiến trúc 2.5. Cách thức lưu trữ và tìm kiếm 2.6. Ví dụ áp dụng		3		3	4	3	
3	Chương 3. Sử dụng Apache Hadoop 3.1. Giới thiệu về mô hình GFS 3.2. Lịch sử Hadoop 3.3. Hadoop trong quản lý và khai thác Big data 3.4. HDFS 3.5. Yarn 3.6. Hadoop I/O		3		3	4	3	
4	Chương 4. Sử dụng Mapreduce 4.1. Giới thiệu 4.2. Các hàm chính 4.3. Hoạt động 4.4. Xây dựng ứng dụng			3	3	4	3	

Chương	Nội dung học phần	Chuẩn đầu ra của học phần						
		CĐR1			CĐR2		CĐR3	
		CĐR 1.1	CĐR 1.2	CĐR 1.3	CĐR 2.1	CĐR 2.2	CĐR 3.1	CĐR 3.2
5	Chương 5. Sử dụng Apache Spark 5.1. Tổng quan 5.2. Các thành phần 5.3. Quản lý bộ nhớ 5.4. Lập trình với RDD 5.5. Xây dựng ứng dụng			3		4	3	4

11. Đánh giá học phần

11.1. Ma trận phương pháp kiểm tra đánh giá với chuẩn đầu ra học phần

STT	Điểm thành phần	Quy định	Trọng số	Phương pháp kiểm tra đánh giá (Hình thức, thời gian, thời điểm)	CĐR của học phần			Ghi chú
					CĐR1	CĐR2	CĐR3	
1	Điểm kiểm tra thường xuyên; điểm đánh giá nhận thức và thái độ; điểm chuyên cần.	01 điểm	20%	+ Hình thức: Vấn đáp; + Thời điểm: Trong các giờ học trên lớp.	CĐR1.1, CĐR1.2, CĐR1.3.	CĐR2.1, CĐR2.2, CĐR2.3.	CĐR3.1, CĐR3.2.	Trung bình cộng các điểm đánh giá
2	Điểm kiểm tra giữa học phần.	01 điểm	30%	+ Hình thức: Thực hành; + Thời gian: 90 phút; + Thời điểm: Giờ học 31, 32 trên lớp.	CĐR1.1, CĐR1.2, CĐR1.3.	CĐR2.1, CĐR2.2, CĐR2.3.	CĐR3.1.	01 bài kiểm tra
3	Điểm thi kết thúc học phần.	01 điểm	50%	+ Hình thức: Bài tập lớn; + Thời gian: 20 phút; + Thời điểm: Theo lịch thi học kỳ.	CĐR1.1, CĐR1.2, CĐR1.3.	CĐR2.1, CĐR2.2, CĐR2.3.	CĐR3.1, CĐR3.2.	01 bài thi

11.2. Cách tính điểm học phần

Điểm học phần là trung bình cộng các điểm thành phần đã nhân trọng số. Tính theo thang điểm 10, làm tròn đến một chữ số thập phân. Sau đó chuyển thành thang điểm chữ và thang điểm điểm 4.

12. Yêu cầu học phần

- Tham gia tối thiểu 80% số tiết học trên lớp dưới sự hướng dẫn của giảng viên.
- Đọc và nghiên cứu tài liệu phục vụ học phần, hoàn thành các bài tập cá nhân và bài tập nhóm.
- Chủ động làm bài tập lớn theo hướng dẫn của giảng viên.
- Tham gia kiểm tra giữa học phần, thi kết thúc học phần.
- Dụng cụ học tập: Máy tính, vở ghi, bút,...

13. Tài liệu phục vụ học phần

- Tài liệu chính:

[1] - Trường Đại học Sao Đỏ (2020), *Giáo trình Dữ liệu lớn - Big data*.

-Tài liệu tham khảo:

[2] - By Krishna Rungta (2019), *Learn Hadoop in 1 Day*.

[3] - Apache HBase™ Reference Guide, *Introduction to Basic Schema Design* by Amandeep Khurana, Version 1.4.11 - ebook.

[4] - Tom White (2015), *Hadoop The Definitive Guide*. Published by O' Reilly Media, Inc., Gravenstein Highway North, Sebastopol, CA 95472.

[5] - Holden Karau Andy Konwinski Matei Zaharia Patrick Wendell (2015), *Learning Spark*. Published by O' Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

14. Nội dung chi tiết học phần và phương pháp dạy-học

TT	Nội dung giảng dạy	Số tiết	Phương pháp dạy-học	CDR học phần
1	Chương 1. Tổng quan về Big data Mục tiêu chương: Sau khi học xong chương này, sinh viên đạt được các yêu cầu cơ bản sau: - Giải thích được khái niệm Big data, các kiểu Big data và đặc trưng của Big data. - Minh họa được các kiểu dữ liệu, các đặc trưng của Big data.	4 (2LT, 2TH)	Thuyết trình; Tổ chức học theo nhóm; Thực hành trên máy tính - Giảng viên: + Giải thích khái niệm, kiểu dữ liệu và đặc trưng của Big data. + Giao bài tập, nội dung thực hành cho cá nhân, các nhóm. + Hướng dẫn sinh viên thực hành, đánh giá, nhận xét.	CDR1.1; CDR3.1.

TT	Nội dung giảng dạy	Số tiết	Phương pháp dạy-học	CDR học phần
	<p>Nội dung cụ thể:</p> <p>1.1. Khái niệm Big data</p> <p>1.1.1. Data</p> <p>1.1.2. Big data</p> <p>1.2. Các kiểu Big data</p> <p>1.2.1. Có cấu trúc</p> <p>1.2.2. Không có cấu trúc</p> <p>1.2.3. Bán cấu trúc</p> <p>1.3. Các đặc trưng của Big data</p> <p>1.3.1. Dung lượng dữ liệu</p> <p>1.3.2. Tốc độ dữ liệu</p> <p>1.3.3. Đa dạng dữ liệu</p> <p>Bài thực hành số 1.</p>		<p>- Sinh viên:</p> <p>+ Đọc trước tài liệu:</p> <p>[1]: Chương 1;</p> <p>[2]: Chương 1;</p> <p>[3]: Chương 2.</p> <p>+ Lắng nghe, ghi chép, quan sát.</p> <p>+ Làm bài tập cá nhân, theo nhóm trong [1]: Chương 1.</p> <p>+ Thực hành bài thực hành số 1.</p>	
2	<p>Chương 2. Sử dụng Hbase</p> <p>Mục tiêu chương:</p> <p>Sau khi học xong chương này, sinh viên đạt được các yêu cầu cơ bản sau:</p> <ul style="list-style-type: none"> - Phân tích được các tính năng của Apache Hbase, mô hình Hbase, kiến trúc và cách lưu trữ dữ liệu của Hbase. - Áp dụng được Hbase lưu trữ dữ liệu trong Big data. <p>Nội dung cụ thể:</p> <p>2.1. Giới thiệu</p> <p>2.2. Các tính năng</p> <p>2.3. Mô hình</p> <p>2.4. Kiến trúc</p> <p>2.5. Lưu trữ và tìm kiếm</p> <p>2.6. Ví dụ áp dụng</p> <p>Bài thực hành số 2 - 4.</p>	12 (6LT, 6TH)	<p>Thuyết trình; Tổ chức học theo nhóm; Thực hành trên máy tính</p> <p>- Giảng viên:</p> <p>+ Giải thích các tính năng và sử dụng của Hbase.</p> <p>+ Nêu nội dung vấn đề cần giải quyết.</p> <p>+ Giao bài tập, nội dung thực hành cho cá nhân và các nhóm.</p> <p>+ Hướng dẫn sinh viên thực hành, đánh giá, nhận xét.</p> <p>- Sinh viên:</p> <p>+ Đọc trước tài liệu:</p> <p>[1]: Chương 2;</p> <p>[3]: Các chương.</p> <p>+ Lắng nghe, ghi chép, quan sát, thảo luận.</p> <p>+ Làm bài tập theo nhóm trong [1]: Chương 2.</p> <p>+ Thực hành bài thực hành số 2 - 4.</p>	CDR1.2; CDR2.1; CDR2.2; CDR3.1.

TT	Nội dung giảng dạy	Số tiết	Phương pháp dạy-học	CDR học phần
3	<p>Chương 3. Sử dụng Apache Hadoop</p> <p>Mục tiêu chương:</p> <p>Sau khi học xong chương này, sinh viên đạt được các yêu cầu cơ bản sau:</p> <ul style="list-style-type: none"> - Phân tích được các tính năng của Apache Hadoop, mô hình Hbase, quản lý và khai thác big data của Hbase. - Đánh giá, lựa chọn được giải pháp quản lý và khai thác dữ liệu trong Big data. <p>Nội dung cụ thể:</p> <p>3.1. Giới thiệu về mô hình GFS</p> <p>3.2. Lịch sử Hadoop</p> <p>3.3. Hadoop trong quản lý và khai thác Big data</p> <p>3.4. HDFS</p> <p>3.5. Yarn</p> <p>3.6. Hadoop I/O</p> <p>Bài thực hành số 5 - 7.</p>	12 (6LT, 6TH)	<p>Thuyết trình; Dạy học dựa trên vấn đề; Tổ chức cho sinh viên thảo luận; Tổ chức học theo nhóm; Thực hành trên máy tính</p> <p>- Giảng viên:</p> <ul style="list-style-type: none"> + Giải thích tính năng, cách sử dụng Hadoop. + Nêu vấn đề, hướng dẫn sinh viên giải quyết vấn đề. + Nêu nội dung thảo luận. + Giao bài tập, nội dung thực hành cho cá nhân, các nhóm. + Hướng dẫn sinh viên thực hành, đánh giá, nhận xét. <p>- Sinh viên:</p> <ul style="list-style-type: none"> + Đọc trước tài liệu: [1]: Chương 3; [4]: Các chương. + Lắng nghe, ghi chép, quan sát, tranh luận, phản biện và giải quyết các vấn đề. + Làm bài tập cá nhân, theo nhóm trong [1]: Chương 3. + Thực hành bài thực hành số 5 - 7. 	CDR1.2; CDR2.1; CDR2.2; CDR3.1.
4	<p>Chương 4. Sử dụng Mapreduce</p> <p>Mục tiêu chương:</p> <p>Sau khi học xong chương này, sinh viên đạt được các yêu cầu cơ bản sau:</p>	12 (6LT, 4TH, 2KT)	<p>Thuyết trình; Dạy học dựa trên vấn đề; Tổ chức học theo nhóm; Thực hành trên máy tính</p> <p>- Giảng viên:</p> <ul style="list-style-type: none"> + Giải thích tính năng và cách sử dụng mô hình Mapreduce. 	CDR1.3; CDR2.1; CDR2.2; CDR3.1.

TT	Nội dung giảng dạy	Số tiết	Phương pháp dạy-học	CDR học phần
	<p>- Phân tích được mô hình Mapreduce, các hàm chính của Mapreduce, hoạt động của Mapreduce, cách thức phát triển Mapreduce.</p> <p>- Đánh giá, lựa chọn được cách sử dụng Mapreduce trong xử lý Big data.</p> <p>Nội dung cụ thể:</p> <p>4.1. Giới thiệu</p> <p>4.2. Các hàm chính</p> <p>4.3. Hoạt động</p> <p>4.4. Xây dựng ứng dụng</p> <p>Kiểm tra giữa học phần</p> <p>Bài thực hành số 8 - 9.</p>		<p>+ Nêu vấn đề, hướng dẫn sinh viên giải quyết vấn đề.</p> <p>+ Giao bài tập, nội dung thực hành cho cá nhân, các nhóm.</p> <p>+ Hướng dẫn sinh viên thực hành, đánh giá, nhận xét.</p> <p>- Sinh viên:</p> <p>+ Đọc trước tài liệu: [1]: Chương 4;</p> <p>+ Lắng nghe, ghi chép, quan sát và giải quyết các vấn đề.</p> <p>+ Làm bài tập cá nhân, theo nhóm trong [1]: Chương 4.</p> <p>+ Làm bài kiểm tra.</p> <p>+ Thực hành bài thực hành số 8 - 9.</p>	
5	<p>Chương 5. Sử dụng Apache Spark</p> <p>Mục tiêu chương:</p> <p>Sau khi học xong chương này, sinh viên đạt được các yêu cầu cơ bản sau:</p> <p>- Phân tích được các thành phần của Apache Spark, các thành phần của Apache Spark, quản lý bộ nhớ và lập trình với RDD.</p> <p>- Đánh giá, lựa chọn được các công cụ vào phân tích xử lý dữ liệu lớn thực tế.</p> <p>Nội dung cụ thể:</p> <p>5.1. Tổng quan</p> <p>5.2. Các thành phần</p>	20 (10LT, 10TH)	<p>Thuyết trình; Dạy học dựa trên vấn đề; Tổ chức học theo nhóm; Thực hành trên máy tính</p> <p>- Giảng viên:</p> <p>+ Giải thích tính năng và cách sử dụng mô hình Spark.</p> <p>+ Nêu vấn đề, hướng dẫn sinh viên giải quyết vấn đề.</p> <p>+ Giao bài tập, nội dung thực hành cho cá nhân, các nhóm.</p> <p>+ Hướng dẫn sinh viên thực hành, đánh giá, nhận xét.</p>	CDR1.3; CDR2.2; CDR3.1; CDR3.2.

TT	Nội dung giảng dạy	Số tiết	Phương pháp dạy-học	CDR học phần
	5.3. Quản lý bộ nhớ 5.4. Lập trình với RDD 5.4.1. Tổng quan 5.4.2. Tạo RDD 5.4.3. Hoạt động của RDD 5.5. Xây dựng ứng dụng Bài thực hành số 10 - 14.		Sinh viên: + Đọc trước tài liệu: [1]: Chương 5; [5]: Các chương. + Lắng nghe, ghi chép, quan sát và giải quyết các vấn đề. + Làm bài tập cá nhân, theo nhóm trong [1]: Chương 5. + Thực hành bài thực hành số 10 - 14.	

Hải Dương, ngày 09 tháng 08 năm 2022

**KT.HIỆU TRƯỞNG
PHÓ HIỆU TRƯỞNG**



TS. Nguyễn Thị Kim Nguyên

**KT. TRƯỞNG KHOA
PHÓ TRƯỞNG KHOA**

Phạm Văn Kiên

TRƯỞNG BỘ MÔN

Vũ Bảo Tạo