



ĐỀ CƯƠNG CHI TIẾT XỬ LÝ DỮ LIỆU LỚN

Mã học phần: BDP34031 – Số tín chỉ: 03

Dùng cho (các) ngành: Công nghệ Thông tin

Điều kiện tiên quyết (nếu có): Các môn học phần cơ sở ngành CNTT

Hình thức đào tạo: Trực tiếp

Đơn vị phụ trách: Khoa Công nghệ Thông tin

1. Mô tả chung về học phần

Học phần giới thiệu tổng quan về khái niệm, đặc trưng cũng như những thách thức của dữ liệu lớn. Giới thiệu một số phương pháp và công cụ phổ biến để khai thác và quản lý dữ liệu lớn (Hadoop, MapReduce và Spark).

2. Các chữ viết tắt (nếu có)

Từ viết tắt	Tiếng anh	Tiếng Việt
CNTT	-	Công nghệ Thông tin
CSDL	-	Cơ sở dữ liệu
HDFS	Hadoop Distributed File System	Hệ thống file lưu trữ của Hadoop

3. Chuẩn đầu ra của học phần

Mã	Chuẩn đầu ra học phần
pl09b	<ol style="list-style-type: none">Sử dụng HDFS để lưu trữ dữ liệu lớn trong môi trường HadoopSử dụng công cụ để phân tích và xử lý dữ liệu lớn.Triển khai ứng dụng big data trong thực tế

4. Tài liệu học tập

4.1. Tài liệu học tập:

Bài giảng môn học – Khoa Công nghệ Thông tin

4.2. Tài liệu tham khảo:

- [1]. Tom White(2015). Hadoop The Definitive Guide. Published by O’ Reilly Media, Inc., Gravenstein Highway North, Sebastopol, CA 95472.
- [2]. David Loshin (2013). Big data analytics. 225 Wyman Street, Waltham, MA 02451, USA
- [3]. Holden Karau, Andy Kowinski and Matei Zaharia(2014). Learning Spark. Published by O’ Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

- [4]. Wes McKinney (2013). Python for data analysis. Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.
- [5]. Jiawei Han, Micheline Kamber, Jian Pei (2012). Data mining Concepts and Techniques. Published by Elsevier, Inc., Waltham, MA 02451, USA.

5. Chiến lược học tập

Sinh viên cần tích cực và chủ động tham gia vào quá trình học tập; cần tham gia đầy đủ các giờ học theo quy định, không ngừng phấn đấu để duy trì sự tiến bộ liên tục trong học tập; hoàn thành nhiệm vụ học tập đúng tiến độ.

Để hoàn thành tốt học phần này, sinh viên cần:

- Tham gia học tập đầy đủ và đúng giờ.
- Đọc tài liệu và làm bài tập trước khi tham dự buổi học kế tiếp.
- Tập trung nghe giảng.
- Thực hành ngay các kiến thức đã học.
- Tích cực thảo luận với giáo viên và bạn học về các nội dung của môn học.

[Mô tả tóm tắt các hoạt động học tập, cách học, kỹ thuật và các gợi ý nhằm giúp sinh viên đạt được các chuẩn đầu ra của học phần]

6. Nội dung, kế hoạch giảng dạy và đánh giá

12	Hoạt động học tập của người học				Chuẩn đầu ra
	Hoạt động trên lớp	ST	Hoạt động tự học	SG	
Giới thiệu môn học Chương 1: Một số thuật toán khai thác dữ liệu 1.1. Phân cụm dữ liệu 1.2. Phân lớp dữ liệu 1.3. Xử lý và phân tích dữ liệu	- Nghe giảng về giới thiệu môn học, Một số thuật toán khai thác dữ liệu - Thảo luận các nội dung liên quan	5	- Đọc trước tài liệu về môn học, Một số thuật toán khai thác dữ liệu - Chuẩn bị câu hỏi, nội dung cần thảo luận	12	plo9b.1 plo9b.2
Chương 2: Tổng quan về dữ liệu lớn 1.1. Khái niệm, đặc trưng về dữ liệu lớn 1.2. Ứng dụng của dữ liệu lớn 1.3. Xử lý song song & phân tán 1.4. Một số công cụ phân tích dữ liệu lớn	- Nghe giảng về Tổng quan về dữ liệu lớn - Thảo luận các nội dung liên quan	4	- Đọc trước tài liệu về Tổng quan về dữ liệu lớn - Chuẩn bị câu hỏi, nội dung cần thảo luận	6	plo9b.1 plo9b.2 plo9b.3

<p>Chương 3: Mô hình quản lý dữ liệu lớn</p> <p>3.1. Giới thiệu Hadoop</p> <p>3.2. Hệ thống file lưu trữ và quản lý của Hadoop Distributed File System</p> <p>3.3. Cơ sở dữ liệu NoSQL</p> <p>3.4. Mô hình MapReduce</p> <p>3.4.1. Giới thiệu về mô hình MapReduce</p> <p>3.4.2. Lưu trữ và phân tích dữ liệu cơ bản với mô hình Hadoop Lưu trữ và phân tích dữ liệu với mô hình Hadoop-NoSQL-MapReduce</p> <p>3.4.3. Lập trình cơ bản trên Hadoop MapReduce</p> <p>3.4.4. Phát triển một ứng dụng Hadoop MapReduce</p> <p>3.4.5. Xây dựng ứng dụng để phân tích dữ liệu lớn trên các tập dữ liệu mẫu có sẵn với mô hình Hadoop-NoSQL-MapReduce</p> <p>3.5. Mô hình Spark</p> <p>3.5.1. Giới thiệu về Apache Spark</p> <p>3.5.2. Lưu trữ và phân tích dữ liệu với mô hình Hadoop Lưu trữ và phân tích dữ liệu với mô hình Hadoop-NoSQL-Spark</p> <p>3.5.3. Lập trình cơ bản trên Hadoop Spark</p> <p>3.5.4. Cách thức phát triển một ứng dụng trên Hadoop Spark</p> <p>3.5.5. Xây dựng ứng dụng để phân tích dữ liệu lớn trên các tập dữ liệu mẫu có sẵn với mô hình Hadoop-NoSQL-Spark.</p>	<ul style="list-style-type: none"> - Nghe giảng về Mô hình quản lý dữ liệu lớn - Thảo luận các nội dung liên quan - Làm bài tập 	21	<ul style="list-style-type: none"> - Đọc trước tài liệu về Mô hình quản lý dữ liệu lớn - Chuẩn bị câu hỏi, nội dung cần thảo luận 	42	<p>plo9b.1</p> <p>plo9b.2</p> <p>plo9b.3</p>
<p>- Thực hành cài đặt và vận hành Hadoop</p>	<p>- Làm bài tập theo hướng dẫn của giảng</p>	30	<p>- Làm bài tập tại nhà theo yêu cầu</p>	30	<p>plo9b.1</p> <p>plo9b.2</p>

- Lưu trữ và phân tích dữ liệu cơ bản với mô hình Hadoop- NoSQL- Spark - Lập trình với Hadoop MapReduce - Lưu trữ và phân tích dữ liệu với mô hình Hadoop-NoSQL-Spark - Lập trình với Hadoop Spark	viên - Thảo luận các nội dung liên quan		của giảng viên - Chuẩn bị câu hỏi, nội dung cần thảo luận		plo9b.3
Đề án môn học Xây dựng ứng dụng để phân tích dữ liệu lớn trên các tập dữ liệu mẫu có sẵn với mô hình Hadoop-NoSQL- MapReduce/Hadoop Spark	- Lựa chọn bài toán - Thu thập dữ liệu, lưu trữ dữ liệu - Phân tích và xử lý dữ liệu - Báo cáo				plo9b.3

ST-Số tiết chuẩn, SG-Số giờ chuẩn.

7. Đánh giá kết quả học tập

Hoạt động đánh giá của học phần gồm:

Phân loại	Phương pháp đánh giá	Tỷ trọng	Chuẩn đầu ra		
			plo9b.1	plo9b.2	plo9b.3
Kết thúc học phần	Bài tập	40%	x	x	
	Báo cáo	60%			x
<i>Tổng cộng:</i>		100%			

7.1. Hoạt động đánh giá - Chuẩn đầu ra: **plo9b.1, plo9b.2** - Tỷ lệ: **40%** điểm học phần

- Hình thức đánh giá: Đánh giá dựa trên các bài tập sinh viên thực hành được giao
- Mô tả bài đánh giá: Bài đánh giá là những bài tập giải quyết từng khía cạnh nhỏ trong môn học. Được giảng viên giao cho sinh viên thực hiện theo tiến độ môn học và nộp cho giảng viên theo thời gian quy định.
- Ma trận đánh giá:

Tiêu chí đánh giá	Khung điểm				
	A	B	C	D	F
	8,5 ÷ 10	7,0 ÷ 8,4	5,5 ÷ 6,9	4,0 ÷ 5,4	< 4,0
TC1: Cài đặt, vận hành, lưu trữ dữ liệu trên Hadoop. Làm bài tập lập trình và xử lý dữ	Cài đặt, vận hành thành thạo Hadoop. Tổ chức dữ liệu, xử lý dữ liệu thể hiện	Cài đặt, vận hành thành thạo Hadoop. Tổ chức dữ liệu, xử lý dữ liệu đạt kết	Cài đặt, vận hành Hadoop. Tổ chức dữ liệu, xử lý dữ liệu đạt kết	Cài đặt, vận hành Hadoop. Tổ chức dữ liệu, xử lý dữ liệu đạt kết	Cài đặt, nhưng không vận hành được Hadoop. Không thực

Tiêu chí đánh giá	Khung điểm				
	A	B	C	D	F
	8,5 ÷ 10	7,0 ÷ 8,4	5,5 ÷ 6,9	4,0 ÷ 5,4	< 4,0
liệu cơ bản trên Hadoop-NoSQL-MapReduce/Hadoop Spark	được tất cả các phần của kết quả phân tích dữ liệu Tương tác tốt với giảng viên và các bạn trong lớp	quả phân tích 60-79% Tương tác tốt với giảng viên và các bạn trong lớp.	quả phân tích 40-59% Ít trao đổi với giáo viên	quả phân tích <40% Thụ động trong tiếp thu kiến thức.	hiện tổ chức dữ liệu, lập trình và xử lý dữ liệu

Kết quả đánh giá chung: CDR1 = 100% TC1

7.1. Hoạt động đánh giá 2 - Chuẩn đầu ra: plo9b.3 - Tỷ lệ: 60% điểm học phần

- Hình thức đánh giá: Báo cáo đồ án môn học
- Mô tả bài đánh giá: Bài đánh giá là một báo cáo trình bày toàn bộ nội dung cuốn báo cáo đồ án môn học. Cuốn báo cáo đồ án môn học là một giải pháp hoàn chỉnh để giải quyết một vấn đề trên thực tế do giảng viên hoặc sinh viên đề xuất được duyệt để thành đồ án môn học.
- Ma trận đánh giá:

Tiêu chí đánh giá	Khung điểm				
	A	B	C	D	F
	8,5 ÷ 10	7,0 ÷ 8,4	5,5 ÷ 6,9	4,0 ÷ 5,4	< 4,0
TC1: Xây dựng ứng dụng để phân tích dữ liệu lớn trên các tập dữ liệu mẫu có sẵn với mô hình Hadoop-NoSQL-MapReduce/Hadoop Spark	Thu thập dữ liệu, lưu trữ dữ liệu, phân tích và xử lý dữ liệu, ứng dụng thể hiện được tất cả các phần của kết quả phân tích dữ liệu	Thu thập dữ liệu, lưu trữ dữ liệu, phân tích và xử lý dữ liệu, kết quả phân tích 60-79%	Thu thập dữ liệu, lưu trữ dữ liệu, phân tích và xử lý dữ liệu, kết quả phân tích 40-59%	Thu thập dữ liệu, lưu trữ dữ liệu, phân tích và xử lý dữ liệu, kết quả đạt được <40%	Thu thập dữ liệu, lưu trữ dữ liệu, Không thực hiện phân tích và xử lý dữ liệu

Kết quả đánh giá chung: CDR2 = 100% TC1

7.2. Cách tính kết quả học tập chung của học phần

Điểm học phần = Điểm CDR1 * 40% + Điểm CDR2 * 60%

8. Các phương tiện, trang thiết bị dạy và học

- Giảng đường, phần, máy chiếu.
- Yêu cầu đối với sinh viên: Có tài liệu môn học, máy tính PC hoặc Laptop

9. An toàn của sinh viên và giảng viên

- Giảng viên và sinh viên phải tuân thủ các quy định về việc sử dụng các trang thiết bị điện tại phòng học.
- Trong trường hợp phát sinh các vấn đề có thể dẫn đến mất an toàn, sinh viên cần kịp thời báo cáo với giảng viên để phối hợp giải quyết.

10. Kỷ luật, khiếu nại và hỗ trợ

- Sinh viên phải có mặt trên lớp đủ thời gian theo quy định của nhà trường
- Sinh viên gian lận trong hoạt động đánh giá nào sẽ hủy kết quả đánh giá đó.
- Sinh viên chưa đạt ĐG nào vẫn tiếp tục học các phần tiếp theo và sẽ được cải thiện điểm trong quá trình học.

Sinh viên gặp bất kỳ khó khăn gì trong quá trình học tập có thể liên hệ trực tiếp với giảng viên, Trưởng khoa/bộ môn, Văn phòng hỗ trợ sinh viên, Phòng Đào tạo, Ban Thanh tra của Nhà trường để được hướng dẫn, hỗ trợ.

**Chủ tịch Hội đồng
xây dựng CTĐT ngành**

Hải Phòng, ngày tháng năm 2022
Người biên soạn

Nguyễn Thị Xuân Hương