|  |  |
| --- | --- |
| TRƯỜNG ĐẠI HỌC VĂN HIẾN | **CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM** |
| **KHOA KỸ THUẬT – CÔNG NGHỆ** | **Độc lập – Tự do – Hạnh phúc** |
|  |  |

**ĐỀ CƯƠNG CHI TIẾT HỌC PHẦN**

1. **Thông tin về học phần**
* Tên học phần: **Phân tích dữ liệu lớn cho viễn thông**
* Mã học phần: **ELE409**
* Số tín chỉ: 3 tín chỉ (2/1/6)
* Bậc đào tạo: Đại học
* Loại học phần (bắt buộc/tự chọn): Tự chọn
* Học phần tiên quyết/ Học phần trước: Cơ sở dữ liệu (INT323)
* Đơn vị phụ trách: Khoa Kỹ thuật – Công nghệ
* Số giờ tín chỉ: 60 tiết, trong đó:
* Lý thuyết: 30 tiết (1 tín chỉ LT = 15 tiết)
* Thực hành: 30 tiết (1 tín chỉ TH = 30 tiết)
1. **Thông tin về giảng viên**

Giảng viên 1:

* Họ và tên: Đàm Quốc Việt
* Chức danh, học vị: Thạc sĩ
* Thời gian làm việc: Giờ hành chính (8:00 -16:00)
* Địa điểm làm việc: Khoa Kỹ thuật – Công nghệ, Trường Đại học Văn Hiến
* Điện thoại: 0376722132
* Email: vietdq@vhu.edu.vn

Giảng viên 2:

* Họ và tên: Hồ Lê Anh Hoàng
* Chức danh, học vị: Thạc sĩ
* Thời gian làm việc: Giờ hành chính (8:00 -16:00)
* Địa điểm làm việc: Khoa Kỹ thuật – Công nghệ, Trường Đại học Văn Hiến
* Điện thoại: 0938117195
* Email: hoanghla@vhu.edu.vn
1. **Tóm tắt nội dung học phần**

﻿Môn học 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).

1. **Mục tiêu của học phần**

Gồm 5 mục tiêu như sau:

* ﻿Sinh viên có khả năng phân tích, xử lý một vấn đề cụ thể liên quan đến Big data.
* ﻿Biết sử dụng Hadoop - HDFS đê lưu trữ dữ liệu lớn.
* ﻿Biết sử dụng mô hình MapReduce và Spark đê phân tích Big data
* ﻿Có khả năng sử dụng ngôn ngữ lập trình Java để phân tích dữ liệu lớn
* ﻿Có khả năng triển khai ứng dụng big data trong thực tế
1. **Chuẩn đầu ra của học phần:**

| **Mã CĐR** | **Nội dung chuẩn đầu ra** |
| --- | --- |
| **Kiến thức** |
| CLO1 | Hiểu ﻿và nắm được các khái niệm cũng như các đặc trưng cơ bản liên quan đến Big data. |
| CLO2 | Có khả năng phân tích, đánh giá các vấn đề liên quan đến Big data trong ngành Điện tử Viễn Thông |
| **Kỹ năng** |
| CLO3 | Có khả năng sử dụng HDFS để lưu trữ dữ liệu lớn trong môi trường Hadoop |
| CLO4 | Có khả năng sử dụng Hadoop-MapReduce để phân tích và xử lý Big data |
| CLO5 | Có khả năng sử dụng Hadoop-Spark để phân tích và xử lý Big data. |
| CLO6 | Có khả năng sử dụng ngôn ngữ lập trình Java để phân tích dữ liệu lớn |
| CLO7 | Có khả năng triển khai ứng dụng big data trong thực tế |
| **Mức độ tự chủ và trách nhiệm** |
| CLO8 | Tổ chức làm việc nhóm, tổng hợp, đánh giá và đề xuất được công việc của các cá nhân trong nhóm thông qua bài tập. |
| CLO9 | Tích cực, chủ động tham gia các hoạt động trên lớp học, tham gia phản biện các nội dung của bài học. |

**Ma trận liên kết giữa Chuẩn đầu ra chương trình đào tạo và Chuẩn đầu ra học phần**

| **Chuẩn đầu ra** | **PLO1** | **PLO2** | **PLO3** | **PLO4** | **PLO5** | **PLO6** | **PLO7** | **PLO8** | **PLO9** | **PLO10** | **PLO11** | **PLO12** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CLO1 | X | X |  | X |  |  |  |  |  |  |  |  |
| CLO2 | X | X |  | X |  |  |  |  |  |  |  |  |
| CLO3 | X | X |  | X | X |  | X |  |  |  |  |  |
| CLO4 | X | X |  | X | X |  | X |  |  |  |  |  |
| CLO5 | X | X |  | X | X |  | X |  |  |  |  |  |
| CLO6 | X | X |  | X | X |  | X |  |  |  |  |  |
| CLO7 | X | X |  | X | X |  | X | X |  |  |  |  |
| CLO8 |  |  |  |  |  |  |  |  |  | X | X |  |
| CLO9 |  |  |  |  |  |  |  |  |  | X | X |  |

***Ghi chú: PLOs*** *(ProgrammeLearningOutcomes): Chuẩn đầu ra cấp CTĐT*

 ***CLOs*** *(Course LearningOutcomes): Chuẩn đầu ra học phần*

1. **Nội dung chi tiết của học phần**

**6.1. Lý thuyết**

| **Chương** | **Nội dung** | **Đáp ứng CLOs** |
| --- | --- | --- |
| **Chương 1** | **Tổng quan về Dữ liệu lớn** | **CLO1, CLO2** |
| 1.1. | ﻿Giới thiệu về khái niệm và một số kỹ thuật khai phá dữ liệu |  |
| 1.2. | Khái niệm về Dữ liệu lớn |  |
| 1.3. | Đặc trưng của Dữ liệu lớn |  |
| 1.4. | Lịch sử hình thành của Dữ liệu lớn |  |
| 1.5. | Thử thách của lĩnh vực Dữ liệu lớn |  |
| **Chương 2** | **Handoop** | **CLO1, CLO2, CLO3, CLO4, CLO5, CLO8, CLO9** |
| 2.1. | Giới thiệu về mô hình GFS(Google File System) và lịch sử ra đời của Hadoop |  |
| 2.2. | ﻿Giải pháp Hadoop cho việc quản lý và khai thác Big data |  |
| 2.3. | Hệ thống file lưu trữ và quản lý của Hadoop: HDFS (Hadoop Distributed FileSystem) |  |
| 2.4. | ﻿ YARN |  |
| 2.5. | Hadoop I/O |  |
| **Chương 3** | **Lập trình MapReduce (MR)** | **CLO1, CLO2, CLO3, CLO4, CLO5, CLO6, CLO7, CLO8, CLO9** |
| 3.1. | ﻿Giới thiệu khả năng dùng Java để viết chương trình xử lý big data (làm việc với MapReduce và Spark). |  |
| 3.2. | ﻿Giới thiệu về mô hình MR |  |
| 3.3. | Cách thức phát triển một ứng dụng MR |  |
| 3.4. | Xây dựng ứng dụng tiêu biểu phân tích Big data trên các tập dữ liệu mẫu có sẵn |  |
| **Chương 4** | **Spark** | **CLO1, CLO2, CLO3, CLO4, CLO5, CLO6, CLO7, CLO8, CLO9** |
| 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 |  |
| 4.4. | ﻿Cách thức phát triển một ứng dụng toàn diện trong việc lưu trữ và phân tích dữ liệu |  |
| 4.5. | ﻿Xây dựng ứng dụng tiêu biểu phân tích Big data trên các tập dữ liệu mẫu có sẵn |  |
| **Chương 5** | **Phân tích dữ liệu lớn** | **CLO1, CLO2, CLO3, CLO4, CLO5, CLO6, CLO7, CLO8, CLO9** |
| 5.1. | Phân tích dữ liệu bảng tính với BigSheets:* ﻿BigSheets là gì?
* Chức năng của BigSheets
* BigSheets chuyên sâu
* Một ví dụ sử dụng BigSheets đầy đủ
 |  |
| 5.2. | Ngôn ngữ R:* Phân tích dữ liệu với ngôn ngữ R
* Sử dụng R với cơ sở dữ liệu
 |  |
| 5.3. | Phân tích dữ liệu thời gian thực |  |

**6.2. Thực hành**

| **Bài** | **Nội dung** | **Đáp ứng CLOs** |
| --- | --- | --- |
| **Bài 1** | **﻿Hadoop** | **CLO1, CLO2, CLO3, CLO4, CLO5, CLO6, CLO7, CLO8, CLO9** |
| 1.1. | Cài đặt Hadoop |  |
| 1.2. | Triển khai 3 chế độ của Hadoop:* ﻿Chế độ Standalone
* Chế độ ﻿Pseudo-Distributed
* Chế độ ﻿Fully-Distributed
 |  |
| 1.3. | Vận hành Hadoop |  |
| **Bài 2** | **Hadoop và MapReduce** | **CLO1, CLO2, CLO3, CLO4, CLO5, CLO6, CLO7, CLO8, CLO9** |
| 2.1. | ﻿Lưu trữ và phân tích dữ liệu cơ bản với mô hình Hadoop – MapReduce |  |
| 2.2. | ﻿﻿Lập trình Java với Hadoop MapReduce |  |
| 2.3. | Triển khai YARN |  |
| **Bài 3** | **Spark** | **CLO1, CLO2, CLO3, CLO4, CLO5, CLO6, CLO7, CLO8, CLO9** |
| 3.1. | ﻿Cài đặt và cấu hình cho Spark |  |
| 3.2. | Lưu trữ và phân tích dữ liệu với mô hình Hadoop – Spark |  |
| 3.3. | ﻿Lập trình Java với Hadoop Spark |  |
| **Bài 4** | **﻿﻿Phân tích dữ liệu với ngôn ngữ R** | **CLO1, CLO2, CLO3, CLO4, CLO5, CLO6, CLO7, CLO8, CLO9** |
| 4.1. | ﻿Cài đặt và cấu hình công cụ cần dùng |  |
| 4.2. | Phân tích dữ liệu nằm trên cơ sở dữ liệu với ngôn ngữ R  |  |

1. **Phân bổ thời gian theo tiết và điều kiện thực hiện**

**7.1. Lý thuyết**

| **Chương** | **Tên chương** | **Số tiết tín chỉ** | **Ghi chú** |
| --- | --- | --- | --- |
| **Lý thuyết** | **Bài tập** | **Thực hành** | **Tự học** | **Tổng** |
| 1 | Tổng quan về Dữ liệu lớn | 3 | 0 | 0 | 5 | 10 |  |
| 2 | Handoop | 7 | 0 | 0 | 10 | 20 |  |
| 3 | Lập trình MapReduce (MR) | 7 | 0 | 0 | 15 | 30 |  |
| 4 | Spark | 6 | 0 | 0 | 15 | 30 |  |
| 5 | Phân tích dữ liệu lớn | 7 | 0 | 0 | 15 | 30 |  |
| **Tổng** | **30** | **0** | **0** | **60** | **90** |  |

**7.2. Thực hành**

| **Bài** | **Tên bài** | **Số tiết tín chỉ** | **Ghi chú** |
| --- | --- | --- | --- |
| **Lý thuyết** | **Bài tập** | **Thực hành** | **Tự học** | **Tổng** |
| 1 | Hadoop | 0 | 0 | 6 | 10 | 25 |  |
| 2 | Hadoop và MapReduce | 0 | 0 | 8 | 16 | 31 |  |
| 3 | Spark | 0 | 0 | 8 | 18 | 33 |  |
| 4 | Phân tích dữ liệu với ngôn ngữ R | 0 | 0 | 8 | 16 | 31 |  |
| **Tổng** | **0** | **0** | **30** | **60** | **90** |  |

**CÁC CHỦ ĐỀ THẢO LUẬN VÀ TIỂU LUẬN**

1.

2.

3.

1. **Phương pháp giảng dạy**

Giảng viên giảng dạy với sự kết hợp của một số phương pháp sau:

* + Thuyết trình
	+ Đọc và tóm lược nội dung tài liệu
	+ Động não nhanh (bài tập tư duy cá nhân)
	+ Thực hành thí nghiệm
	+ Trình bày trực quan
	+ Giao bài đọc về nhà
	+ Hướng dẫn tự học
	+ Thảo luận nhóm

**Ma trận liên kết giữa Chuẩn đầu ra với phương pháp giảng dạy**

| **Phương pháp giảng dạy** | **CLO1** | **CLO2** | **CLO3** | **CLO4** | **CLO5** | **CLO6** | **CLO7** | **CLO8** | **CLO9** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Thuyết trình | X | X | X | X | X | X | X | X | X |
| Đọc và tóm lược nội dung tài liệu | X | X | X | X | X | X | X | X | X |
| Động não nhanh (bài tập tư duy cá nhân) | X | X | X | X | X | X | X | X | X |
| Thực hành thí nghiệm | X | X | X | X | X | X | X | X | X |
| Trình bày trực quan | X | X | X | X | X | X | X | X | X |
| Giao bài đọc về nhà | X | X | X | X | X | X | X | X | X |
| Hướng dẫn tự học | X | X | X | X | X | X | X | X | X |
| Thảo luận nhóm | X | X | X | X | X | X | X | X | X |

1. **Phương pháp học tập**

Sinh viên học tập với sự kết hợp của một số phương pháp sau:

* + Thuyết trình
	+ Làm việc nhóm
	+ Tự học, tự nghiên cứu
	+ Làm thí nghiệm theo nhóm
	+ Tìm kiếm thông tin/tài liệu

**Ma trận liên kết giữa Chuẩn đầu ra với phương pháp học tập**

| **Phương pháp học tập** | **CLO1** | **CLO2** | **CLO3** | **CLO4** | **CLO5** | **CLO6** | **CLO7** | **CLO8** | **CLO9** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Thuyết trình | X | X | X | X | X | X | X | X | X |
| Làm việc nhóm | X | X | X | X | X | X | X | X | X |
| Tự học, tự nghiên cứu | X | X | X | X | X | X | X | X | X |
| Làm thí nghiệm theo nhóm | X | X | X | X | X | X | X | X | X |
| Tìm kiếm thông tin/tài liệu | X | X | X | X | X | X | X | X | X |

1. **Nhiệm vụ của sinh viên**
	1. - Chủ động tổ chức thực hiện giờ tự học.
	2. - Tham gia đầy đủ các giờ lên lớp và giờ thuyết trình (sinh viên chỉ được vắng mặt tối đa 20% thời gian lên lớp của học phần).
	3. - Đọc tài liệu tham khảo bắt buộc và bổ trợ do giảng viên giới thiệu.
	4. - Hoàn thành đầy đủ các bài tập cá nhân, bài tập nhóm.
	5. - Tham gia kỳ thi kết thúc học phần.
2. **Thang điểm đánh giá:** Điểm đánh giá quá trình và điểm thi kết thúc học phần được chấm theo thang điểm 10 (từ 0 đến 10), làm tròn đến 1 chữ số thập phân.
3. **Phương pháp kiểm tra, đánh giá kết quả học tập**

Sinh viên được đánh giá kết quả học tập trên cơ sở hai điểm thành phần như sau:

1. *Điểm đánh giá quá trình: trọng số 40% bao gồm:*
	1. Điểm chuyên cần: trọng số 10%
	2. Điểm kiểm tra giữa kỳ: trọng số 30%
2. *Điểm thi kết thúc học phần: trọng số 60%*

Hình thức thi: Tự luận viết/Báo cáo thực hành

**Ma trận quan hệ giữa Chuẩn đầu ra và phương pháp kiểm tra, đánh giá**

| **Hình thức đánh giá** | **CLO1** | **CLO2** | **CLO3** | **CLO4** | **CLO5** | **CLO6** | **CLO7** | **CLO8** | **CLO9** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Báo cáo thực hành | X | X | X | X | X | X | X | X | X |
| Tự luận viết | X | X | X | X | X | X | X | X | X |
| Thuyết trình | X | X | X | X | X | X | X | X | X |
| Dự lớp |  |  |  |  |  |  |  | X | X |

1. **Tài liệu phục vụ cho học phần**

**13.1. Tài liệu chính**

- Hossam Abou El-Nasr, Aisha A. N. Mohamed, (2018), Big Data for Telecommunications: Challenges and Opportunities, Springer.

J. S. H. Choi, (2020), Data Mining for the Internet of Things: How to Mine Data from IoT Devices and Use it for Business Insights, Springer.

**13.2. Tài liệu tham khảo**

- Bernard Marr, (2024), Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results, Ascent Audio.

*TP.Hồ Chí Minh, ngày 15 tháng 04 năm 2025*

|  |  |  |
| --- | --- | --- |
| **HIỆU TRƯỞNG****PGS.TS. Nguyễn Minh Đức** | **Trưởng khoa/Bộ môn****TS. Đinh Thị Thủy** | **Giảng viên biên soạn****ThS. Đàm Quốc Việt** |