|  |  |
| --- | --- |
| BỘ GIÁO DỤC VÀ ĐÀO TẠO  **TRƯỜNG ĐẠI HỌC**  **KINH DOANH VÀ CÔNG NGHỆ HÀ NỘI**  **KHOA CÔNG NGHỆ THÔNG TIN** | **CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM**  **Độc lập – Tự do – Hạnh phúc**  *Hà Nội,ngày tháng năm 20* |

**ĐỀ CƯƠNG CHI TIẾT HỌC PHẦN**

**BIG DATA**

**1. Thông tin chung về học phần**

***1.1. Tên học phần: Big data , Mã số học phần:*** 191032706

***1.2. Số tín chỉ: 2***

***1.3. Thuộc chương trình đào tạo trình độ: Đại học , hình thức đào tạo:  
Chính qui***

***1.4. Đơn vị thực hiện (Khoa, Bộ môn): Mạng và cơ sở dữ liệu***

***1.5 Loại học phần: bắt buộc***

***1.6. Điều kiện tiên quyết: Không có***

***1.7. Giờ tín chỉ đối với các hoạt động:***

- Nghe giảng lý thuyết : 15 tiết

- Thực hành : 30 tiết

- Tự học : 120 tiết

**2. Mục tiêu của học phần**

***2.1. Kiến thức***

* Hiểu các chức năng cốt lõi trong Big data và biết tích hợp chúng tạo thành các giải pháp áp dụng thực tiễn.
* Xác định thách thức chính của Big Data trong một số lãnh vực bao gồm cả phương tiện truyền thông xã hội, giao thông vận tải, tài chính, y học…

***2.2. Kỹ năng***

* **Tìm hiểu và áp dụng Big Data vào thực tiễn.**
* **Tìm hiểu làm thế nào để tối đa hóa các mô hình lập trình ứng dụng vào các hệ thống dữ liệu lớn.**

***2.3. Thái độ***

**- Rèn luyện tác phong làm việc khoa học, chuẩn xác.**

**-Có khả năng hoạt động nhóm tốt.**

**- Thường xuyên cập nhật, bắt kịp xu thê công nghệ mới,** những kết quả nghiên cứu mới**.**

**3. Tóm tắt nội dung học phần**

Cung cấp cho sinh viên kiến thức tổng quan về các khái niệm, các kỹ thuật, các công cụ và công nghệ liên quan đến Big Data; các mô hình lưu trữ khác nhau, các phương pháp xử lý và các công cụ báo cáo có sẵn để làm việc với Big data; các chức năng cốt lõi của mỗi thành phần trong Big data và tích hợp để tạo thành giải pháp thống nhất mang lại lợi ích.

**4.** **Nội dung chi tiết học phần**

|  |  |  |  |
| --- | --- | --- | --- |
| **Nội dung** | **Hình thức hoạt động dạy và học** | | |
| Lên lớp | | Sinh viên tự nghiên cứu, tự học |
| Lý thuyết | Thực hành |
| **Chương 1: Khoa học dữ liệu, BigData & thách thức**  1. Khoa học dữ liệu và thách thức thời đại  2. Dữ liệu lớn  3. Công nghệ xử lý dữ liệu mới | 3 | 2 | 15 |
| **Chương 2: Thu thập Big Data**  1. Đặc trưng dữ liệu lớn và thu thập dữ liệu  2.  Các công cụ và giao thức để tích hợp dữ liệu  3.  Vấn đề làm sạch, biến đổi, và củng cố dữ liệu | 2 | 3 | 25 |
| **Chương 3: Lưu trữ Big Data**  1.  Cơ sở dữ liệu hiện đại  2.   Giải pháp quản lý dữ liệu ngày nay, bao gồm cả RDBMS truyền thống, NoSQL, DFS.  3.  Hệ thống tính toán song song cho phép phân phối xử lý dữ liệu trên cluster | 4 | 6 | 20 |
| **Chương 4: Hệ thống Big Data và Cloud**  1. Tổng quan về Dữ liệu lớn trên Đám mây.  2. Nền tảng đám mây hỗ trợ dữ liệu lớn.  -   Thảo luận về hiệu suất, khả năng mở rộng, và chi phí khi lưu trữ dữ liệu trên đám mây | 2 | 3 | 15 |
| **Chương 5: Phân tích Big Data**  1.Descriptive analysis (Phân tích miêu tả)  2.Diagnostic Analysis (Phân tích chẩn đoán)  3.Predictive Analysis (Phân tích dự đoán)  4.Prescriptive Analysis (Phân tích đề xuất) | 2 | 3 | 25 |
| **Chương 6: Nền tảng công nghệ phân tích Big Data**   * 1. Apache Hadoop   2. Apache Spark   3. Kiến trúc khác | 2 | 3 | 20 |
| **Thực hành khởi nghiệp** |  | 10 |  |
| **Tổng cộng** | 15 | 30 | 120 |

**5. Tài liệu học tập**

***5.1. Tài liệu chính***

Giáo trình Big Data - Trường Đại học Kinh doanh và Công nghệ Hà Nội – Lưu hành nội bộ

***5.2. Tài liệu tham khảo***

[1] Paul Zikopoulos, Dirk deRoos, David Corrigan, Tom Deutsch, Krishnan Parasuraman, James Giles, *Harness the Power of Big Data*, Mc Graw Hill, 2012.

[2] Donald Miner, Adam Shook, *MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems*, O'Reilly Media, 2012.

[3] Tom White, Doug Cutting, *Hadoop: The Definitive Guide*, O'Reilly Media, 2015.

[4] Judith Hurwitz, Alan Nugent, Dr. Fern Halper, Marcia Kaufman, *Big Data for Dummies*, John Wiley & Sons. Inc, 2013.

[5] Dirk deRoos, Paul C. Zikopoulos, Roman B. Melnyk, Bruce Brown, Rafael Coss, *Hadoop for Dummies*, John Wiley & Sons. Inc, 2014.

**6. Hướng dẫn giảng viên thực hiện và yêu cầu đối với sinh viên**

|  |  |  |
| --- | --- | --- |
| **Nội dung** | **Nhiệm vụ của giảng viên** | **Nhiệm vụ của sinh viên** |
| **Chương 1: Khoa học dữ liệu, BigData & thách thức** |  |  |
| 1. Khoa học dữ liệu và thách thức thời đại  2. Dữ liệu lớn | Giảng viên làm nổi bật khái niệm, tầm quan trọng Bigdata |  |
| 3. Công nghệ xử lý dữ liệu mới Thu thập (acquire)  Tổ chức (organize) Phân tích (analyze)  Quyết định (decide) | Nêu ra khái quát các vấn đề ứng dụng, sử dụng Dữ liệu lớn |  |
|  |  |  |
| Nội dung thực hành | - Giảng viên giao bài tập  - Hướng dẫn cách thực hiện |  |
| **Chương 2: Thu thập Big Data** |  |  |
| 1. Đặc trưng dữ liệu lớn và thu thập dữ liệu  2.  Các công cụ và giao thức để tích hợp dữ liệu | Xử lý làm sạch dữ liệu vì ban đầu, dữ liệu thô |  |
| 3. Vấn đề làm sạch, biến đổi, và củng cố dữ liệu |  |  |
| Nội dung thực hành | - Giảng viên giao bài tập  - Hướng dẫn cách thực hiện  - Kiểm tra, đánh giá kết quả thực hiện bài tập của sinh viên |  |
| **Chương 3: Lưu trữ Big Data** |  |  |
| 1.  Cơ sở dữ liệu hiện đại  2.   Giải pháp quản lý dữ liệu ngày nay, bao gồm cả RDBMS truyền thống, NoSQL, DFS. | Giảng viên hướng dẫn phân tích ưu nhược điểm của các giải pháp |  |
| 3.  Hệ thống tính toán song song cho phép phân phối xử lý dữ liệu trên cluster | Giảng viên giới thiệu về hệ thống tính toán, xử lý song song |  |
| Nội dung thực hành | - Giảng viên giao bài tập  - Hướng dẫn cách thực hiện  - Kiểm tra, đánh giá kết quả thực hiện bài tập của sinh viên |  |
| **Chương 4: Hệ thống Big Data** |  |  |
| 1. Tổng quan về Dữ liệu lớn trên Đám mây. | Khái quát hóa mối quan hệ giữa Cloud và Big data, nền tảng của điện toán đám mây sẽ bổ sung cho việc xử lý, lưu trữ và phân tích dữ liệu của kiến trúc big data |  |
| 2. Nền tảng đám mây hỗ trợ dữ liệu lớn.  -   Thảo luận về hiệu suất, khả năng mở rộng, và chi phí khi lưu trữ dữ liệu trên đám mây | Google  Microsoft  Amazon  Cloudera |  |
| Nội dung thực hành | - Giảng viên giao bài tập  - Hướng dẫn cách thực hiện |  |
| **Chương 5: Phân tích Big Data** |  |  |
| 1.Descriptive analysis (Phân tích miêu tả)  2.Diagnostic Analysis (Phân tích chẩn đoán)  3.Predictive Analysis (Phân tích dự đoán)  4.Prescriptive Analysis (Phân tích đề xuất) | Hướng dẫn sinh viên về các phương thức phân tích Bigdata |  |
| **Chương 6: Nền tảng công nghệ phân tích Big Data** |  |  |
| 1 Apache Hadoop  2 Apache Spark 3. Kiến trúc khác | Hướng dẫn sinh viên về các nền tảng phân tích |  |
| Nội dung thực hành | - Giảng viên giao bài tập  - Hướng dẫn cách thực hiện  - Kiểm tra, đánh giá kết quả thực hiện bài tập của sinh viên | Thực hiện theo yêu cầu của giảng viên |

**7. Phương pháp, hình thức kiểm tra - đánh giá kết quả học tập học phần**

**7.1. Thang điểm đánh giá**

Giảng viên đánh giá theo thang điểm 10

**7.2. Kiểm tra đánh giá quá trình**

Có trọng số tối đa 40%, bao gồm điểm đánh giá bộ phận như sau:

* Điểm chuyên cần: 10%
* Nhận thức, thái độ học tập trên lớp, tham gia thảo luận, semina, bài tập: 10%
* Bài kiểm tra giữa kỳ: 20%

**7.3. Điểm thi kết thúc học phần**

Điểm thi kết thúc học phần có trọng số 60%

Hình thức thi: Trắc nghiệm

**8.Thông tin về giảng viên xây dựng đề cương chi tiết HP:**

*Họ tên: Phan Bảo Định*

*Học hàm, học vị :Thạc sỹ*

**9. Phê duyệt của Khoa**

|  |  |  |
| --- | --- | --- |
| **Chủ nhiệm Khoa** |  | **Người biên soạn**  Ths.Phan Bảo Định |