|  |  |
| --- | --- |
| TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HCM  **KHOA CÔNG NGHỆ THÔNG TIN** |  |

ĐỀ CƯƠNG CHI TIẾT HỌC PHẦN

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1. Tên học phần | | Tên tiếng Việt: **Mã nguồn mở trong Khoa học dữ liệu**  Tên tiếng Anh: **Open Source tools for Data Science** | | | | | | | | | | | | | | | | | | | | | | |
| 1. Mã học phần | | CMP1044 | | | | | | | | | | | | | | | | | | | | | | |
| 1. Thuộc khối kiến thức | | Kiến thức GD đại cương | | | | | | | | | | | Kiến thức GD chuyên nghiệp | | | | | | | | | | | |
| 🗖 Kiến thức bắt buộc | | | 🗖 Kiến thức tự chọn | | | | | | | | ⌧ Kiến thức bắt buộc | | | | | | | 🗖 Kiến thức tự chọn | | | | |
| 1. Trình độ đào tạo | | Đại học | | | | | | | | | | | | | | | | | | | | | | |
| 1. Số tín chỉ | | **3 (3,0) TC** | | | | | | | | | | | | | | | | | | | | | | |
| 1. Học phần học trước/ song hành | |  | | | | | | | | | | | | | | | | | | | | | | |
| 1. Mục tiêu của học phần | | Học phần cung cấp cho sinh viên kiến thức và kĩ năng để xử lý dữ liệu trước khi bước vào việc tìm những thông tin có ý nghĩa của dữ liệu. Sinh viên biết cách thu thập dữ liệu, khai phá dữ liệu: tổng hợp, làm sạch, chuẩn hóa, biến đổi dữ liệu kinh doanh, thống kê tạo ra các báo cáo thông qua việc sử dụng các bộ thư viện, công cụ mã nguồn mở mạnh mẽ. | | | | | | | | | | | | | | | | | | | | | | |
| 1. Đơn vị quản lý học phần | | Khoa Công nghệ Thông tin | | | | | | | | | | | | | | | | | | | | | | |
| 1. **Bảng trích ngang ma trận sự đóng góp của mỗi học phần cho chuẩn đầu ra của CTĐT** | | | | | | | | | | | | | | | | | | | | | | | | |
| Mã HP | Học phần | | | Chuẩn đầu ra CTĐT (PLO) | | | | | | | | | | | | | | | | | | | | |
| PLO1 | | | PLO2 | | | PLO3 | | PLO4 | | | PLO5 | | PLO6 | PLO7 | | | | PLO8 | | PLO9 |
| CMP1044 | **Mã nguồn mở trong Khoa học dữ liệu** | | | - | | | PI2.1/2  PI2.2/3 | | | PI3.1/3  PI3.2/3 | | - | | | PI5.3/2 | | - | PI7.3/2 | | | | - | | - |
| 1. Chuẩn đầu ra của học phần (CLO) | | | | | | | | | | | | | | | | | | | | | | | | |
| **Chuẩn đầu ra học phần**  ***(Course Learning Outcome)*** | | | | | | | | | | | | | | | | **Chỉ số đo lường (PI) việc đạt CĐR của CTĐT** | | | | | **Tương ứng với CĐR của CTĐT** | | | |
| * CLO1: Trình bày được các kiến thức nền tảng về các mã nguồn mở trong Khoa học dữ liệu. * CLO2: Mô tả và giải thích được sự khác biệt giữa cơ sở dữ liệu quan hệ và NoSQL. Vận dụng vào chuẩn bị, thăm dò và khai thác dữ liệu cho từng loại cơ sở dữ liệu NoSQL. * CLO3: Thiết kế các các cơ sở dữ liệu tài liệu, cột rộng và biểu đồ. * CLO4: Có ý thức kỷ luật, có tinh thần trách nhiệm cao, tác phong chuyên nghiệp và tự tin trong công việc. | | | | | | | | | | | | | | | | PI2.1  PI2.1, PI2.2  PI3.2  PI3.1  PI3.2  PI5.3, PI7.3 | | | | | PLO2  PLO2  PLO3  PLO3  PLO3  PLO5, PLO7 | | | |
| 1. **Mô tả tóm tắt nội dung học phần**   Cơ sở dữ liệu NoSQL để lưu trữ lượng lớn dữ liệu phức tạp. Tìm hiểu về sự khác biệt giữa cơ sở dữ liệu quan hệ và NoSQL, xem xét các loại cơ sở dữ liệu NoSQL và xem cách thực hiện các nhiệm vụ khoa học dữ liệu chung, bao gồm chuẩn bị dữ liệu, thăm dò và xây dựng và áp dụng mô hình.  Nội dung môn học bắt đầu với phần giới thiệu về NoSQL và sau đó nghiên cứu kỹ các cơ sở dữ liệu tài liệu, cột rộng và biểu đồ. Tìm hiểu các chi tiết chính để thực hiện chuẩn bị, thăm dò và khai thác dữ liệu cho từng loại cơ sở dữ liệu NoSQL, cách sử dụng các cơ sở dữ liệu NoSQL khác nhau với các công cụ khoa học dữ liệu phổ biến, bao gồm cơ sở dữ liệu tài liệu MongoDB, cơ sở dữ liệu cột rộng Cassandra và cơ sở dữ liệu biểu đồ Neo4j. | | | | | | | | | | | | | | | | | | | | | | | | |
| 1. Phương pháp, hình thức tổ chức dạy học của học phần | | Thuyết giảng (Lecture): Cung cấp cho sinh viên hệ thống kiến thức nền tảng của môn học một cách khoa học, logic.  Câu hỏi gợi mở (Inquiry): Thông qua việc hỏi đáp giữa giáo viên và sinh viên để làm rõ các nội dung kiến thức trong môn học.  Mô hình (Models): Giúp sinh viên hiểu rõ và biết vận dụng các nội dung môn học vào các vấn đề thực tiễn.  Học nhóm (Pear Learning): Giúp người học tăng cường năng lực tự học, tự nghiên cứu. | | | | | | | | | | | | | | | | | | | | | | |
| 1. **Nội dung chi tiết học phần** | | | | | | | | | | | | | | | | | | | | | | | | |
| **BÀI SỐ**  *(1)* | | TÊN BÀI *(2)* | | | | | | **Số tiết** | | | **Phương pháp, hình thức**  **tổ chức dạy học**  *(tham khảo phụ lục 2)*  *(5)* | | | | | | | | | | | | **Đáp ứng CĐR**  **của HP**  *(5)* | |
| **LT***(3)* | **TH** *(4)* | |
| **BÀI 1** | | Giới thiệu NoSQL | | | | | | **6** |  | | - GV: Giới thiệu các khái niệm về NoSQL và Các loại cơ sở dữ liệu NoSQL  - SV: Hiểu được vai trò, ứng dụng của NoSQL | | | | | | | | | | | | CLO1 | |
|  | | Các giới hạn của cơ sở dữ liệu quan hệ | | | | | |  |  | |
| 1.2. | | Các loại cơ sở dữ liệu NoSQL | | | | | |  |  | |
| 1.3. | | Ưu điểm của cơ sở dữ liệu NoSQL | | | | | |  |  | |
| **Bài 2** | | Thực hiện các nhiệm vụ khoa học dữ liệu với NoSQL | | | | | | **9** |  | | - GV: Chuẩn bị dữ liệu, khám phá dữ liệu, xây dựng mô hình và áp dụng mô hình  - SV: Hiểu được vai trò, thực hiện các nhiệm vụ khoa học dữ liệu với NoSQL | | | | | | | | | | | | CLO1 CLO2 | |
| 2.1. | | Chuẩn bị dữ liệu | | | | | |  |  | |
| 2.2. | | Khám phá dữ liệu | | | | | |  |  | |
| 2.3. | | Xây dựng mô hình | | | | | |  |  | |
| 2.4. | | Áp dụng mô hình | | | | | |  |  | |
| **Bài 3** | | Document Databases for Data Science | | | | | | **12** |  | | - GV: Giới thiệu mô hình dữ liệu tài liệu, giới thiệu khung dữ liệu trong MongoDB  - SV: Hiểu được vai trò, sử dụng cơ sở dữ liệu tài liệu cho khoa học dữ liệu | | | | | | | | | | | | CLO1 CLO3 | |
| 3.1. | | Mô hình dữ liệu tài liệu | | | | | |  |  | |
| 3.2. | | Cấu trúc JSON | | | | | |  |  | |
| 3.3. | | Chuẩn bị dữ liệu với cơ sở dữ liệu tài liệu | | | | | |  |  | |
| 3.4. | | Cài đặt Anaconda | | | | | |  |  | |
| 3.5. | | Cài đặt MongoDB | | | | | |  |  | |
| 3.6. | | Làm việc với Jupyter | | | | | |  |  | |
| 3.7. | | Khám phá dữ liệu với cơ sở dữ liệu tài liệu | | | | | |  |  | |
| 3.8. | | Trích xuất dữ liệu với cơ sở dữ liệu tài liệu | | | | | |  |  | |
| 3.9. | | Thực hiện kiểm tra chất lượng | | | | | |  |  | |
| 3.10. | | Lập chỉ mục dữ liệu với cơ sở dữ liệu tài liệu | | | | | |  |  | |
| 3.11 | | Khung dữ liệu trong MongoDB | | | | | |  |  | |
| 3.12 | | Sử dụng cơ sở dữ liệu tài liệu cho khoa học dữ liệu | | | | | |  |  | |
| **Bài 4** | | Wide-Column Databases for Data Science | | | | | | **9** |  | | - GV: Hướng dẫn thực thi mô hình dữ liệu cột rộng và chuẩn bị dữ liệu với cơ sở dữ liệu cột rộng  - SV: Cài đặt Cassandra và Chuẩn bị dữ liệu cho Cassandra | | | | | | | | | | | | CLO2 | |
| 4.1 | | Mô hình dữ liệu cột rộng | | | | | |  |  | |
| 4.2 | | Chuẩn bị dữ liệu với cơ sở dữ liệu cột rộng | | | | | |  |  | |
| 4.3 | | Cài đặt Bộ phát triển Java | | | | | |  |  | |
| 4.4 | | Cài đặt Cassandra | | | | | |  |  | |
| 4.5 | | Chuẩn bị dữ liệu cho Cassandra | | | | | |  |  | |
| 4.6 | | Tải dữ liệu vào Cassandra | | | | | |  |  | |
| 4.7 | | Cassandra và Spark | | | | | |  |  | |
| **Bài 5** | | Graph Databases for Data Science | | | | | | **9** |  | | - GV: Giới thiệu và hướng dẫn khái niệm graphi chính và chuẩn bị dữ liệu với cơ sở dữ liệu biểu đồ  - SV: Hiểu và có thể sử dụng khám phá dữ liệu, trích xuất dữ liệu với cơ sở dữ liệu biểu đồ | | | | | | | | | | | | CLO2  CLO4 | |
| 5.1 | | Mô hình dữ liệu đồ thị | | | | | |  |  | |
| 5.2 | | Khái niệm graphi chính | | | | | |  |  | |
| 5.3 | | Chuẩn bị dữ liệu với cơ sở dữ liệu biểu đồ | | | | | |  |  | |
| 5.4 | | Cài đặt Neo4j | | | | | |  |  | |
| 5.6 | | Khám phá dữ liệu với cơ sở dữ liệu biểu đồ | | | | | |  |  | |
| 5.7 | | Trích xuất dữ liệu với cơ sở dữ liệu biểu đồ | | | | | |  |  | |
| **TỔNG CỘNG:** | | | | | | | | **45** |  | |  | | | | | | | | | | | |  | |
| 1. **Phương pháp kiểm tra/đánh giá** | | Điểm thành phần | Quy định | | | | | | | | | | | Bài đánh giá | | | | | **Trọng số** | | | | **Đáp ứng CĐR của HP** | |
| Điểm đánh giá quá trình | - Chuyên cần  - Điểm bài tập về nhà  - Điểm kiểm tra | | | | | | | | | | | Bài 1 - 5 | | | | | 15%  15%  20%  50% | | | | CLO1  CLO2  CLO3  CLO4 | |
| Điểm thi kết thúc HP | Đồ án môn học | | | | | | | | | | | Bài 1,3,5 | | | | | 50% | | | | CLO2  CLO4 | |
| 1. **Tài liệu phục vụ học phần** | | Tài liệu/giáo trình chính | * Khoa Công nghệ Thông tin. Tài liệu học tập học phần “Mã nguồn mở trong Khoa học dữ liệu”. HUTECH | | | | | | | | | | | | | | | | | | | | | |
| Tài liệu tham khảo/bổ sung | * MongoDB The Definitive Guide: Powerful and Scalable Data Storage (Kindle Edition) * SQL & NoSQL Databases: Models, Languages, Consistency Options and Architectures for Big Data Management (Kindle Edition) * Graph Databases 2nd Edition, By Ian Robinson, Jim Webber, and Emil Eifrém | | | | | | | | | | | | | | | | | | | | | |
| Các phần mềm | * MongoDB, Cassandra và Neo4j | | | | | | | | | | | | | | | | | | | | | |
| 1. **Hướng dẫn sinh viên tự học** | | **Nội dung** | | | | **Số tiết** | | | **Nhiệm vụ của sinh viên** | | | | | | | | | | | | | | | |
| **Bài 1: Giới thiệu NoSQL**  1.1 Các giới hạn của cơ sở dữ liệu quan hệ  1.2 Các loại cơ sở dữ liệu NoSQL  1.3 Ưu điểm của cơ sở dữ liệu NoSQL | | | | **20** | | | Làm bài tập và báo cáo | | | | | | | | | | | | | | | |
| **Bài 2: Thực hiện các nhiệm vụ khoa học dữ liệu với NoSQL**  2.1 Chuẩn bị dữ liệu  2.2 Khám phá dữ liệu  2.3 Xây dựng mô hình  2.4 Áp dụng mô hình | | | | **20** | | | Làm bài tập và báo cáo | | | | | | | | | | | | | | | |
| **Bài 3: Document Databases for Data Science**  3.1 Mô hình dữ liệu tài liệu  3.2 Cấu trúc JSON  3.3 Chuẩn bị dữ liệu với cơ sở dữ liệu tài liệu  3.4 Cài đặt Anaconda  3.5 Cài đặt MongoDB  3.6 Làm việc với Jupyter  3.7 Khám phá dữ liệu với cơ sở dữ liệu tài liệu  3.8 Trích xuất dữ liệu với cơ sở dữ liệu tài liệu  3.9 Thực hiện kiểm tra chất lượng  3.10 Lập chỉ mục dữ liệu với cơ sở dữ liệu tài liệu  3.11 Khung dữ liệu trong MongoDB  3.12 Sử dụng cơ sở dữ liệu tài liệu cho khoa học dữ liệu | | | | **20** | | | Làm bài tập và báo cáo | | | | | | | | | | | | | | | |
| **Bài 4: Wide-Column Databases for Data Science**  4.1 Mô hình dữ liệu cột rộng  4.2 Chuẩn bị dữ liệu với cơ sở dữ liệu cột rộng  4.3 Cài đặt Bộ phát triển Java  4.4 Cài đặt Cassandra  4.5 Chuẩn bị dữ liệu cho Cassandra  4.6 Tải dữ liệu vào Cassandra  4.7 Cassandra và Spark | | | | **20** | | | Làm bài tập và báo cáo | | | | | | | | | | | | | | | |
| **Bài 5: Graph Databases for Data Science**  5.1 Mô hình dữ liệu đồ thị  5.2 Khái niệm graphi chính  5.3 Chuẩn bị dữ liệu với cơ sở dữ liệu biểu đồ  5.4 Cài đặt Neo4j | | | | **20** | | | Làm bài tập và báo cáo | | | | | | | | | | | | | | | |
| 1. **Các yêu cầu của HP** | | ……………………………………………………………………………………….. | | | | | | | | | | | | | | | | | | | | | | |

*TP. Hồ Chí Minh, ngày… tháng…năm 2022*

**Trưởng khoa/viện Trưởng ngành/bộ môn Người biên soạn**

**\* Ghi chú tổng quát:**

*Trường hợp đề cương học phần cần được cung cấp cho sinh viên thì bổ sung thêm thông tin sau:*

**Giảng viên phụ trách học phần**

|  |  |
| --- | --- |
| Họ và tên: Bùi Mạnh Toàn | Học hàm, học vị: Thạc sĩ |
| Địa chỉ cơ quan:  Trường đại học Công nghệ Tp.HCM – 475A Điện Biên Phủ, Phường 25, Quận Bình Thạnh, Tp.HCM | Điện thoại liên hệ: 0367485420 |
| Email: bm.toan@hutech.edu.vn | Trang web: |

**Giảng viên hỗ trợ học phần/trợ giảng (nếu có)**

|  |  |
| --- | --- |
| Họ và tên: | Học hàm, học vị: |
| Địa chỉ cơ quan: | Điện thoại liên hệ: |
| Email: | Trang web: |

|  |  |
| --- | --- |
| Cách liên lạc với giảng viên: | *(nêu rõ hình thức liên lạc giữa sinh viên với giảng viên/trợ giảng)* |

🙠🕮🙢