|  |  |
| --- | --- |
| **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: **Thực hành mã nguồn mở trong Khoa học dữ liệu**  Tên tiếng Anh: **Practice of Open Source tools for Data Science** | | | | | | | | | | | | | | | | | | |
| 1. Mã học phần | | CMP3052 | | | | | | | | | | | | | | | | | | |
| 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 | | **1 (0 , 1) TC** | | | | | | | | | | | | | | | | | | |
| 1. Học phần học trước/ song hành | |  | | | | | | | | | | | | | | | | | | |
| 1. Mục tiêu của học phần | | Học phần cung cấp cho sinh viên các 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. Từ đó có thể làm việc độc lập hoặc nhóm dự án công nghệ thông tin, có thể đảm nhận vị trí kĩ sư phát triển phần mềm trong doanh nghiệp. | | | | | | | | | | | | | | | | | | |
| 1. Đơn vị quản lý HP | | 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 |
| CMP3052 | Thực hành mã nguồn mở trong Khoa học dữ liệu | | | - | PI2.1/2  PI2.2/3 | PI3.1/3  PI3.2/3 | | - | | | PI5.3/2 | | - | | - | | - | | | - |
| 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: Mô tả và giải thích được sự khác biệt giữa cơ sở dữ liệu quan hệ và NoSQL. * CLO2: Thiết kế các cơ sở dữ liệu tài liệu, cột rộng và biểu đồ * CLO3: Thăm dò và khai thác dữ liệu cho từng loại cơ sở dữ liệu NoSQL * CLO4: Tuân thủ giấy phép phần mềm trong việc sử dụng, thay đổi và tái phân phối phần mềm mã nguồn mở | | | | | | | PI2.1  PI2.2, PI3.1  PI3.2  PI5.3 | | | | | | | PLO2  PLO2, PLO3  PLO3  PLO5 | | | | | | |
| 1. **Mô tả tóm tắt nội dung học phần**   Nội dung học phần 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 | | | **Phương pháp** | | | **Mục tiêu** | | | | | | | | | | | | | | |
| Thuyết giảng | | | Cung cấp cho sinh viên hệ thống kiến thức nền tảng của học phần một cách khoa học, logic. | | | | | | | | | | | | | | |
| Bài tập thực hành | | | Giúp sinh viên hiểu rõ và biết vận dụng các nội dung học phần vào các vấn đề thực tiễn. | | | | | | | | | | | | | | |
| Nghiên cứu bài học, đọc tài liệu tham khảo | | | Giúp sinh viên chủ động trong việc tự nghiên cứu các kiến thức cần thiết. Nâng cao ý thức và trách nhiệm tự học, tự bồi dưỡng kiến thức. | | | | | | | | | | | | | | |
| 1. **Nội dung chi tiết học phần** | | | | | | | | | | | | | | | | | | | | |
| **BÀI SỐ** | | TÊN BÀI | | | | | **Số tiết** | | | | | | | **Phương pháp, hình thức tổ chức dạy học** | | | | **Đáp ứng CĐR của HP**  *(ghi số tương ứng)* | | |
| **LT** | | | **TH** | |  | | | | | | |  | | |
| **Bài 1** | | Giới thiệu NoSQL | | | | |  | | | **5** | | Thuyết giảng: Giới thiệu về NoSQL và sự khác biệt với kiểu dữ liệu truyền thống  Bài tập thực hành: Làm báo cáo về các loại CSDL | | | | | | | CLO1 | | |
| **Bài 2** | | Thực hiện các nhiệm vụ khoa học dữ liệu với NoSQL | | | | |  | | | **5** | | Thuyết giảng: Các bước chuẩn bị để khám phá dữ liệu  Bài tập thực hành: Làm báo cáo về các loại các bước chuẩn bị và khám phá dữ liêj | | | | | | | CLO3 | | |
| **Bài 3** | | Document Databases for Data Science | | | | |  | | | **5** | | Thuyết giảng: Giới thiệu về Json và các công cụ khai thác Json  Bài tập thực hành: Cài đặt các công cụ để thao tác với kiểu dữ liệu Json và làm các hoạt động để khai thác dữ liệu Json | | | | | | | CLO3 | | |
| **Bài 4** | | Wide-Column Databases for Data Science | | | | |  | | | **5** | | Thuyết giảng: Giới thiệu về dữ liệu cột rộng và các công cụ thao tác với dữ liêu cột rộng  Bài tập thực hành: Cài đặt các công cụ để thao tác với kiểu dữ liệu cột rộng và làm các hoạt động để khai thác dữ liệu cột rộng | | | | | | | CLO2 | | |
| **Bài 5** | | Graph Databases for Data Science | | | | |  | | | **5** | | Thuyết giảng: Chuẩn bị dữ liệu cho Graph và cài đặt Neo4j  Bài tập thực hành: Chuẩn bị dữ liệu cho Graph và cài đặt Neo4j | | | | | | | CLO3  CLO4 | | |
| **Bài 6** | | Graph Databases for Data Science (cont) | | | | |  | | | 5 | | Thuyết giảng: Thao thác khai thác với dữ liệu Graph và Neo4j  Bài tập thực hành: Thao thác khai thác với dữ liệu Graph và Neo4j | | | | | | | CLO3  CLO4 | | |
| **TỔNG CỘNG:** | | | | | | |  | | | **30** | |  | | | | | | |  | | |
| 1. **Phương pháp đánh giá** | | Điểm thành phần | | | Quy định | | Bài đánh giá | | | | | | | **Trọng số** | | | | **Đáp ứng CĐR của HP** | | |
| Điểm đánh giá môn học | | | Chuyên cần, tích cực | | | |  | | | | | 30% | | | |  | | |
| Trung bình điểm các buổi thực hành. | | | | Bài 1 - 6 | | | | | 70% | | | | CLO1  CLO2  CLO3  CLO4 | | |
| 1. **Tài liệu phục vụ học phần** | | Tài liệu/giáo trình chính | | | * Khoa Công nghệ Thông tin. Tài liệu học tập học phần   “Thực hành 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 | | | | | **12** | | | 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 | | | | | **18** | | | 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 | | | | | **24** | | | 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 | | | | | **18** | | | 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 | | | | | **18** | | | Làm bài tập và báo cáo | | | | | | | | | | | |

*Tp. Hồ Chí Minh, ngày… tháng…năm …..*

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