

Trường Đại học Bách Khoa -
ĐHQG Tp.HCM
Khoa: Khoa Khoa học và Kỹ
thuật Máy tính
Khoa/Bộ môn quản lý MH: Hệ
thống Thông tin

Tp.HCM, ngày tháng
năm

Đề cương môn học Sau đại học

KỸ THUẬT DỮ LIỆU (DATA ENGINEERING)

Mã số MH: CO5240

Số tín chỉ:	Tc (LT,BT&TH.Tự Học): 3						ECTS: 6					
Số tiết	-Tổng:	70.5	LT:	30	BT:	9	TH:	0	ĐA:		BTL/TL:	31.5
Đánh giá:	Thuyết trình:				30%	Thuyết trình một số nội dung môn học						
	Tiểu luận:				30%	Thảo luận với GV để chọn đề tài, làm việc theo nhóm (2SV/nhóm)						
	Cuối kỳ:				40%	Trắc nghiệm + tự luận						
- Môn tiên quyết:												
- Môn học trước:												
- Môn song hành:												
- CTĐT ngành (Mã ngành): Khoa Học Máy Tính (8480101)												
- Ghi chú khác:												

1. Mục tiêu môn học:

Môn học này cung cấp và củng cố cho học viên những kiến thức và kỹ năng trong lĩnh vực kỹ thuật dữ liệu (data engineering):

- Phân tích, thiết kế cơ sở dữ liệu theo các mô hình dữ liệu khác nhau như mô hình quan hệ, phi cấu trúc (NoSQL)
- Các vấn đề về lập chỉ mục, xử lý giao tác, xử lý linh hoạt và tối ưu hóa câu truy vấn
- Các kiến thức liên quan đến cơ sở dữ liệu hướng đối tượng, hướng thời gian, dữ liệu lớn và các ứng dụng.
- Những chủ đề nghiên cứu nâng cao về cơ sở dữ liệu và các ứng dụng mới cũng sẽ được giới thiệu và thảo luận.

Aims:

The subject is to revise and provide knowledge and skills in data engineering:

- Database analysis and design methodologies including relational data model and NoSQL models
- Advances in data storage and retrieval methods, flexible query answering, query processing and optimization, transaction processing
- Object-oriented and time series databases, as well as big data and applications.
- Research topics relevant to modern databases and applications will also be introduced and discussed.

2. Nội dung tóm tắt môn học:

Nội dung cơ bản của môn học gồm các phần chính sau: Thiết kế ý niệm cơ sở dữ liệu; Mô hình dữ liệu quan hệ; Thiết kế cơ sở dữ liệu mức logic và vật lý; Lập chỉ mục trong cơ sở dữ liệu; Xử lý uyển chuyển câu truy vấn; Xử lý giao tác; Xử lý và tối ưu hóa câu truy vấn; Kho dữ liệu; Cơ sở dữ liệu hướng đối tượng, hướng thời gian; Dữ liệu lớn và ứng dụng; Những hướng nghiên cứu mới.

Course outline:

The course outline is presented as follows: Revision of conceptual and logical/physical database design; Advanced indexing techniques and flexible query answering in databases; Object and object-relational databases; Query processing and optimization; Transaction processing, concurrency control techniques and database recovery; ; Data warehousing; Spatial, temporal, and spatio-temporal databases; Introduction to Big data and applications; and Emerging research directions.

3. Tài liệu học tập:

- [1] R. Elmasri, S.B. Navathe: “*Fundamentals of Database Systems*”, 7th Edition, Pearson Addison-Wesley, 2016
- [2] P.A. Bernstein, E. Newcomer: “*Principles of Transaction Processing*”, 2nd Edition, Elsevier Inc., 2009
- [3] S. Lightstone, T. Teorey, T. Nadeau: “*Physical Database Design*”, Elsevier Inc., 2007
- [4] R. Kimball, M. Ross, “*The Data Warehouse ToolKit*”, 3rd Edition, *Wiley Publishing, Inc.*, 2013.
- [5] J. Hurwitz, A. Nugent, F. Halper, M. Kaufman: “*Big Data for Dummies*”, John Wiley & Sons Inc., 2013
- [6] Jiawei Han, Micheline Kamber, Jian Pei, “*Data Mining: Concepts and Techniques*”, Third Edition, Morgan Kaufmann Publishers, 2012.
- [7] Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman: *Mining of Massive Datasets*, 2014.

4. Các hiểu biết, các kỹ năng cần đạt được sau khi học môn học:

STT	Chuẩn đầu ra môn học (CĐRMH)	Công cụ đánh giá CĐRMH	Đóng góp CĐR Chương trình (CĐRCT)		
			Ứng dụng	Nghiên cứu	
CĐRMH.1	Nắm được các nguyên lý về phân tích và thiết kế CSDL (thiết kế ý niệm, luận lý, vật lý)	Thuyết trình, Cuối kỳ	a		1.3, 1.3
CĐRMH.2	Có kiến thức và kỹ năng về lập chỉ mục, xử lý giao tác, xử lý linh hoạt và tối ưu hóa câu truy vấn	Thuyết trình, Tiểu luận	e		2.1

	Chuẩn đầu ra của chương trình (CĐRCT)										
Chuẩn đầu ra môn học (CĐRMH)	a	b	c	d	e	f	g	h	i	j	k
CĐRMH.5											

5. Hướng dẫn cách học - chi tiết cách đánh giá môn học:

Học viên phải:

- Tham dự giờ giảng trên lớp + làm bài tập đầy đủ
- Thuyết trình: 30% (thuyết trình các nội dung bài học)
- Thực hiện tiểu luận: 30%, làm việc nhóm và thuyết trình trên lớp
- Thi cuối kỳ: 40%, viết và trắc nghiệm
- Ghi chú về điều kiện cấm thi, cách tổng kết điểm: học viên không được thiếu bất cứ cột điểm nào của môn học mới được tính điểm sau cùng của môn; trừ 1 điểm tổng kết khi vắng 3 buổi, trừ tiếp một điểm khi vắng buổi tiếp theo.

Sinh viên phải:

- có đầy đủ các cột điểm
- tham gia đầy đủ các buổi học: vắng 3 buổi sẽ bị trừ 1 điểm trong điểm tổng, mỗi buổi tiếp theo sẽ bị trừ thêm 1 điểm.

Learning strategies & Assessment Scheme:

Student should:

- Attend the class and finish exercises as required
- Presentation: 30% (present topics)
- Project: 30% (team work and presentation)
- Attend the final exam: 40%
- Note: Students must finish all components of the subject and take all the required exams, presentations, etc.; -1 point from the total score if absence 3 lectures, minus further 1 for each absent lecture.

Students must:

- obtain all score columns
- attend all lectures: -1 point in the total score if absence 3 lectures, -1 more point for each additional absence.

6. Nội dung chi tiết:

Tuần/ Buổi	Chủ đề (chương)	Nội dung	Chuẩn đầu ra môn học	Tài liệu
1	Giới thiệu môn học Ôn tập	Giới thiệu môn học Ôn tập: Khái niệm và kiến trúc hệ cơ sở dữ liệu Thiết kế CSDL mức ý niệm: Mô hình thực thể mối liên kết và mở rộng Bài tập/kiểm tra	CĐRMH.1	[1]
2	Các tinh chỉnh trong thiết kế dữ liệu	Phân tích sự phù hợp của hệ thống CSDL đã thiết kế (cân nhắc về dạng chuẩn và tính linh động, hiệu suất của hệ thống) Các nguyên lý tinh chỉnh trong thiết kế vật lý CSDL	CĐRMH.1	[1]
3	Lập chỉ mục nâng cao trong CSDL và tăng hiệu quả truy vấn	Cấu trúc dữ liệu và kỹ thuật lập chỉ mục Các kỹ thuật để tăng hiệu quả truy vấn	CĐRMH.2	Tất cả
4 - 5	Thảo luận về các mô hình dữ liệu phi quan hệ	Thảo luận về các mô hình dữ liệu phi quan hệ: XML, JSON, NoSQL,... Ưu/nhược điểm, hướng sử dụng của các hệ thống NoSQL (HBase, Redis, Elasticsearch, Cassandra, MongoDB,...)	CĐRMH.4	Tất cả
6	Xử lý dữ liệu theo gói và theo dòng	Xử lý dữ liệu theo gói và theo dòng Hệ sinh thái Hadoop, Spark, Hive	CĐRMH.4, CĐRMH.5	Tất cả
7-8	Giới thiệu dữ liệu lớn và ứng dụng	Giới thiệu dữ liệu lớn và ứng dụng Các công cụ quản lý dữ liệu lớn Các công cụ tổ chức và xử lý dữ liệu của Google (BigQuery, BigTable, Tensorflow,...) IoTs và Big Data	CĐRMH.4, CĐRMH.5	[1-4], Internet
9-10	Kho dữ liệu	Các khái niệm về kho dữ liệu Các nguyên lý và công cụ thiết kế, triển khai kho dữ liệu	CĐRMH.3	Tất cả
11-13	Bài tập tổng quan tại lớp	Bài tập tổng quan tại lớp Báo cáo tiểu luận	CĐRMH.1, CĐRMH.2, CĐRMH.3, CĐRMH.4, CĐRMH.5	Tất cả
(Không g lên lớp)	PHẦN BÀI TẬP, TIỂU LUẬN (32 tiết): Sinh viên thực hiện và báo cáo tiểu luận với GV. GV không lên lớp cho các buổi tiểu luận, thay vào đó HV báo cáo với GV tại phòng Lab của GV.	- HV thảo luận với GV trong việc chọn đề tài, mục tiêu và cách tiếp cận- HV báo cáo tiến độ làm việc với GV và tiếp nhận góp ý, định hướng của GV- HV báo cáo kết quả tiểu luận và GV đánh giá (tuần 11-13)	CĐRMH.5	Tất cả

7. Giảng viên tham gia giảng dạy:

CBGD
chính:

PGS.TS
Trần
Minh
Quang

CBGD
tham
gia:

PGS.TS
Võ Thị
Ngọc
Châu
TS.
Phan
Trọng
Nhân

**XÁC NHẬN
CỦA HỘI
ĐỒNG XÂY
DỰNG
CHƯƠNG
TRÌNH ĐÀO
TẠO VÀ KHOA**

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

.....
**GIẢNG
VIÊN
LẬP ĐỀ
CƯƠNG**

**PGS.TS
Phạm
Hoàng
Anh**