

ĐỀ CƯƠNG MÔN HỌC
IS217 – KHO DỮ LIỆU VÀ OLAP

1. THÔNG TIN CHUNG

Tên môn học (tiếng Việt):	Kho dữ liệu và OLAP
Tên môn học (tiếng Anh):	Data Warehouse and OLAP
Mã môn học:	IS217
Thuộc khối kiến thức:	Chuyên ngành
Khoa/Bộ môn phụ trách:	Hệ thống thông tin
Website môn học	
Giảng viên phụ trách:	ThS. Đỗ Thị Minh Phụng Email: phungdtm@uit.edu.vn
Số tín chỉ:	3 TC lý thuyết: 3 TC thực hành: 0
Tính chất của môn	Bắt buộc ngành HTTT
Điều kiện đăng ký: (<i>môn học trước</i>)	Cơ sở dữ liệu

2. MÔ TẢ MÔN HỌC

Sinh viên được trang bị kiến thức về kho dữ liệu và các phương pháp phân tích, thiết kế kho dữ liệu, các mô hình dữ liệu đa chiều, ngôn ngữ truy vấn cơ sở dữ liệu đa chiều để xây dựng các ứng dụng thực tế cho doanh nghiệp, và các kỹ năng mô phỏng CSDL dạng khối, kỹ năng phân tích dữ liệu đa chiều, khai phá dữ liệu, kỹ năng trích xuất, biến đổi và nạp dữ liệu vào kho, vận dụng công cụ BI thành thạo và ngôn ngữ truy vấn dữ liệu đa chiều.

3. MỤC TIÊU MÔN HỌC

Lý thuyết

Mục tiêu	Mô tả [1]	Mục tiêu (Theo CDR cấp 3) [2]
G1	Nắm được các khái niệm căn bản kho dữ liệu, đặc tính, kiến trúc kho dữ liệu, phân loại kho dữ liệu, lược đồ kho dữ liệu, khai phá dữ liệu và kỹ thuật khai phá dữ liệu cơ bản	1.3.10
G2	Có khả năng phân tích nhu cầu báo cáo trực tuyến phục vụ các doanh nghiệp hay các loại hình tổ chức khác.	4.3.1
G3	Xác định chức năng, thành phần và kiến trúc kho dữ liệu	4.3.2
G4	Vận dụng quy trình thiết kế và phát triển kho dữ liệu	4.4.1
G5	Sử dụng được công cụ tích hợp dữ liệu (SSIS) và phân tích dữ liệu (SSAS), công cụ báo biểu (SSRS), hỗ trợ ra quyết định (DSS)	4.4.3
G6	Nắm vững kỹ năng mô phỏng CSDL dạng khối, vận dụng thành thạo ngôn ngữ truy vấn dữ liệu đa chiều để đề xuất các giải pháp và ước lượng.	2.1.3
G7	Dựa trên kết quả ước lượng, chọn giải pháp thiết kế, phân tích kho dữ liệu phù hợp nhất và khuyến nghị	2.1.4

4. CHUẨN ĐẦU RA MÔN HỌC

CĐRMH [1]	CDR cấp 4 của CTĐT [2]	Mô tả CĐRMH (mục tiêu cụ thể) [3]	Mức độ giảng dạy [4]
G1.1	1.3.10.1	Nắm vững các khái niệm căn bản kho dữ liệu, đặc tính, kiến trúc kho dữ liệu, phân loại kho dữ liệu, lược đồ kho dữ liệu	TU
G1.2	1.3.10.2	Nắm được khái niệm khai phá dữ liệu và kỹ thuật khai phá dữ liệu cơ bản	TU
G2.1	4.3.1.1	Phân tích nhu cầu báo cáo trực tuyến phục vụ các doanh nghiệp hay các loại hình tổ chức khác.	TU

G3.1	4.3.2.1	Phân tích và xác định chức năng, kiến trúc kho dữ liệu	TU
G3.2	4.3.2.2	Xác định các phương án lược đồ kho dữ liệu, công nghệ xử lý liên quan	TU
G4.1	4.4.1.1	Lựa chọn nguồn dữ liệu dựa trên mục tiêu phân tích	TU
G4.2	4.4.1.2	Phân tích các phương án lược đồ trong thiết kế	TU
G4.3	4.4.1.3 4.4.1.4	Đánh giá và lựa chọn lược đồ ưu tiên	TU
G5.1	4.4.3.1	Sử dụng các công cụ BI như: tích hợp dữ liệu (SSIS) và phân tích dữ liệu (SSAS), công cụ báo biểu (SSRS), hỗ trợ ra quyết định (DSS)	TU
G6.1	2.1.3.1	Mô phỏng CSDL dạng khối, truy vấn dữ liệu đa chiều, đề xuất các giải pháp thiết kế kho, xử lý phân tích dữ liệu đa chiều	TU
G6.2	2.1.3.2	Ước lượng kết quả thiết kế, phân tích	T
G7.1	2.1.4.1 2.1.4.2 2.1.4.3	Phân tích kết quả các giải pháp và kiểm tra, đánh giá được chất lượng của một kho dữ liệu, đề xuất và khuyến nghị	TU

5. NỘI DUNG MÔN HỌC, KẾ HOẠCH GIẢNG DẠY

Tuần	Nội dung giảng dạy	CDR môn học	Hoạt động dạy và học	Thành phần đánh giá
1	Chương 1: TỔNG QUAN VÀ KIẾN TRÚC KHO DỮ LIỆU 1.1 Nhu cầu doanh nghiệp 1.2 Khái niệm và mục đích xây dựng kho dữ liệu 1.3 Đặc tính của kho dữ liệu 1.4 Các thành phần kho dữ liệu 1.5 Cấu trúc dữ liệu (Dữ liệu chi tiết hiện hành, chi tiết cũ, tổng hợp sơ bộ, mức cao, siêu dữ liệu) 1.6 Dòng dữ liệu trong kho dữ liệu 1.7 Các ứng dụng kho dữ liệu	G1.1 G2.1	Dạy: giải thích nhu cầu sử dụng dữ liệu trong doanh nghiệp Học ở lớp: khảo sát tài liệu liên quan, sử dụng công cụ tìm kiếm online	A3
2,3	1.6 Kiến trúc kho dữ liệu (theo	G1.1	Dạy: trình bày các	

	DATAMART, theo tổng thể mức doanh nghiệp). 1.7 So sánh kho dữ liệu và hệ thống CSDL tác nghiệp 1.8 Siêu dữ liệu		khái niệm cơ bản và đặt yêu cầu so sánh kho dữ liệu và hệ thống csdl tác nghiệp Học ở lớp: khảo sát tài liệu liên quan và trả lời câu hỏi giảng viên	
4	Chương 2: PHÂN TÍCH VÀ THIẾT KẾ KHO DỮ LIỆU 2.1 Lược đồ kho dữ liệu: lược đồ hình sao, lược đồ bông tuyết, CSDL đa chiều 2.2 Các phương pháp thiết kế từ trên xuống, từ dưới lên, xác định các bảng chiều và sự kiện, xây dựng các độ đo,...	G1.1 G3.1 G3.2 G4.2 G4.3	Dạy: trình bày các khái niệm về lược đồ hình sao, bông tuyết, khái niệm bảng sự kiện, bảng chiều, các phương pháp thiết kế, đặt câu hỏi và cho bài tập trên lớp (slides) Học ở lớp: nghe giảng, vận dụng kiến thức đã học để trả lời câu hỏi, làm bài tập trên lớp Học ở nhà: bài tập 1	A1.1 A3
5	2.3 Mô hình dữ liệu luận lý 2.4 Mô hình dữ liệu vật lý 2.5 Vấn đề hiệu suất	G4.2 G4.3	Dạy: trình bày các mô hình dữ liệu Học ở nhà: bài tập 1 (tiếp tục).	A1.1
6	Chương 3: XÂY DỰNG VÀ QUẢN LÝ KHO DỮ LIỆU 3.1 Các nguồn dữ liệu 3.2 Trích xuất dữ liệu 3.3 Biến đổi dữ liệu 3.4 Tải/Nạp dữ liệu vào kho dữ liệu - Demo sử dụng công cụ BI cho quá trình SSIS	G4.1 G4.2 G5.1	Dạy: trình bày phương pháp trích xuất dữ liệu, phương pháp biến đổi, nạp dữ liệu, minh họa bằng phần mềm SSIS Học ở nhà: bài tập 2	A1.2
7	3.5 Duy trì dữ liệu trong kho dữ liệu 3.6 Phát triển kho dữ liệu	G4.3	Dạy: trình bày sự phát triển kho dữ liệu Học ở nhà: bài tập 2 (tiếp theo)	A1.2

8, 9	Chương 4: PHÂN TÍCH DỮ LIỆU TRỰC TUYẾN (OLAP) 4.1 Mô phỏng CSDL dạng khối và phân tích dữ liệu (OLAP, ROLAP & MOLAP) - Demo sử dụng công cụ BI cho quá trình SSAS - Demo sử dụng công cụ BI cho quá trình SSRS	G5.1 G6.1 G6.2	Dạy: trình bày các pp mô phỏng csdl dạng khối, phân tích dữ liệu đa chiều, thực hiện các report, minh họa bằng phần mềm BI với 2 công cụ SSAS, SSRS	A3
10, 11	4.2 Ngôn ngữ truy vấn đa chiều	G6.1 G6.2	Dạy: trình bày ngôn ngữ truy vấn đa chiều, minh họa bằng ví dụ, phần mềm Học ở lớp: làm bài tập trên lớp qua case study Học ở nhà: bài tập 3	A1.3 A3
12	Demo triển khai xây dựng một kho dữ liệu mẫu theo chủ đề, xử lý trích xuất, biến đổi, nạp dữ liệu và phân tích dữ liệu đa chiều, thực hiện một số báo cáo mẫu.	G4.1 G6.1 G6.2 G7.1	Dạy: triển khai xây dựng kho dữ liệu mẫu Adventure Works hoặc Book Store,...đặt câu hỏi. Học ở lớp: sinh viên thảo luận, trả lời câu hỏi, tự đánh giá lại bài tập của mình	
13	Chương 5: HỖ TRỢ RA QUYẾT ĐỊNH (DSS) 5.1 Giới thiệu các phương pháp khai phá dữ liệu (data mining)	G1.2	Dạy: Thuyết giảng, thảo luận và đặt câu hỏi cho sinh viên. Học ở lớp: sinh viên vận dụng các kiến thức đã học để trả lời câu hỏi, thảo luận.	A3
14	5.2 Ứng dụng công cụ BI hỗ trợ ra quyết định - Xác định kỹ thuật khai phá dữ liệu.	G1.2 G5.1	Dạy: demo công cụ BI hỗ trợ ra quyết định, và đặt câu hỏi cho sinh viên.	

	- Cấu trúc khai phá (input, output, các tham số cho dữ liệu huấn luyện và dữ liệu thử,...)		Học ở lớp: sinh viên vận dụng các kiến thức đã học để trả lời câu hỏi, thảo luận, tham gia xây dựng bài học.	
15	Góp ý bài tập nhóm (đồ án) Ôn tập		Dạy: - Góp ý, nhận xét, trên các đồ án của nhóm sinh viên (trước đó, giảng viên chỉ định một vài nhóm để seminar) - Nêu các hiểu nhầm thường gặp (common errors) - Học ở lớp: Sinh viên đặt câu hỏi	

6. PHƯƠNG PHÁP GIẢNG DẠY VÀ HỌC TẬP

- GV lên kế hoạch học tập và cung cấp các tài liệu cần thiết, SV dành nhiều thời gian để chủ động trong việc tự học và tự tìm hiểu thêm các tài liệu liên quan dưới sự hướng dẫn của GV.
- Thực hiện các bài tập về nhà, đồ án môn học (1-2 sinh viên).
- Sinh viên vắng quá 3 buổi học trên lớp và không thực hiện đồ án môn học sẽ không được tham dự thi lý thuyết cuối kỳ.
- Hình thức thi cuối kỳ: tự luận và trắc nghiệm

7. HÌNH THỨC ĐÁNH GIÁ KẾT QUẢ HỌC TẬP

- Đồ án môn học dựa trên một bài toán thực tế được thực hiện theo nhóm từ 1 đến 2 sinh viên, đi từ phân tích, thiết kế kho dữ liệu đến cài đặt cụ thể và khai thác với phần mềm SQL Server Business Intelligence Development Studio. Một số công cụ hỗ trợ khác cũng được áp dụng cho đồ án môn học.
- Hình thành nhóm, nhóm thảo luận, phân công công việc và lập bảng kế hoạch thực hiện để các thành viên nhóm theo dõi.

Thành phần đánh giá [1]	Nội dung [2]	CĐRMH (Gx) [3]	Tỷ lệ % [4]
I). Điểm quá trình gồm Bài tập về nhà (A1.1 (2%), A1.2 (3%), A1.3 (5%) + ĐỒ ÁN MÔN HỌC (40%))	<p>Triển khai xây dựng kho dữ liệu từ cơ sở dữ liệu mẫu Adventure Works hoặc Book Store,.. và xử lý phân tích dữ liệu bằng công cụ BI và SQL Server DBMS.</p> <p>- A1.1: Phân tích nhu cầu báo cáo và đề xuất thiết kế một kho dữ liệu từ cơ sở dữ liệu mẫu Adventure Works hoặc Book Store,...theo chủ đề tùy chọn.</p> <p>- A1.2: Thiết kế quá trình trích xuất dữ liệu, biến đổi, nạp dữ liệu vào kho dữ liệu ở bài tập 1 (Quá trình SSIS)</p> <p>- A1.3: Xử lý phân tích dữ liệu trực tuyến từ kho, hỗ trợ ra quyết định (Quá trình SSAS và SSRS)</p> <p>- Đồ án Môn học (40%): Phân tích, thiết kế hoàn chỉnh một ứng dụng kho dữ liệu trong doanh nghiệp và hỗ trợ ra quyết định (Xem quy định thực hiện đồ án và các mức điểm thành phần).</p>	G2, G3, G4, G5, G6, G7	50%
II). Lý thuyết cuối kỳ	Tự luận, trắc nghiệm	G1, G2, G3, G6	50%

8. TÀI LIỆU HỌC TẬP, THAM KHẢO

Tài liệu học tập:

[1]. W.H. Inmon, Building the Data Warehouses”, Willey Dreamtech, 2004.

[2]. Slide bài giảng lưu hành nội bộ - Khoa HTTT.

Tham khảo:

[3]. E.G. Mallach, “Decision Support and Data Warehouse systems”, 2001.

[4]. Marco Frailis, Data Management and Mining in Astrophysical Databases, 2003.

[5]. Paulraj Ponniah, “Data Warehousing Fundamentals”, John Willey, 2003

[6]. Ralph Kimball, “The Data Warehouse Lifecycle toolkit”, John Willey, 2002

[7]. R. Kimball, “The Data Warehouse Toolkit”, John Willey, 2004.

9. PHẦN MỀM HAY CÔNG CỤ HỖ TRỢ THỰC HÀNH

1. SQL Server 2008, trong đó có cài đặt thêm chức năng SQL Server Business Intelligence Development Studio được tích hợp vào trong Visual Studio 2008.
2. MS Excel.