

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
Giảng viên biên soạn:	ThS. Đỗ Thị Minh Phụng, ThS. Nguyễn Thị Kim Phụng, Email: phungdtm@uit.edu.vn , phungntk@uit.edu.vn
Số tín chỉ:	3
	TC lý thuyết: 3 TC thực hành: 0
Môn học trước:	Cơ sở dữ liệu

2. MÔ TẢ MÔN HỌC (Course description)

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 (Course Goals)

Bảng 1.

Mục tiêu	Mục tiêu môn học	Chuẩn đầu ra trong CTĐT
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.	LO2 (2.7)
G2	Vận dụng kỹ năng xác định vấn đề, hình thành ý tưởng, chọn giải pháp thiết kế, phân tích kho dữ liệu phù hợp nhất và khuyến nghị.	LO4 (4.1, 4.2)

4. CHUẨN ĐẦU RA MÔN HỌC (Course learning outcomes)

Bảng 2. (NT: Nhận thức, KN: Kỹ năng, TĐ: Thái độ)

CĐRMH [1]	LOs [2]	Mô tả CĐRMH [3]	Cấp độ CĐR Môn học [4]
G1.1	2.7	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, khai phá dữ liệu và kỹ thuật khai phá dữ liệu cơ bản.	NT3
G2.1	4.1	- Vận dụng kỹ năng phân tích bài toán thực tế : + Phân tích nhu cầu báo cáo trực tuyến phục vụ các doanh nghiệp hay các loại hình tổ chức khác. + Phân tích và xác định chức năng, kiến trúc kho dữ liệu. + Xác định các phương án lược đồ kho dữ liệu, công nghệ xử lý liên quan. + Lựa chọn nguồn dữ liệu dựa trên mục tiêu phân tích. + Phân tích các phương án lược đồ trong thiết kế. + Đánh giá và lựa chọn lược đồ ưu tiên.	KN4
G2.2	4.1 4.2	- Vận dụng kỹ năng sử dụng các công cụ BI, mô hình hóa, thiết kế nhằm giải quyết bài toán thực tế : + Sử dụng các công cụ BI như: tích hợp dữ liệu (SSIS) và phân tích dữ liệu (SSAS, Power BI, Excel), hỗ trợ ra quyết định (DSS).	KN4

		+ 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. + Ước lượng kết quả thiết kế, phân tích. + 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.	
--	--	---	--

5. NỘI DUNG CHI TIẾT

Tuần/ Thời lượng	Nội dung	Ghi chú/Mô tả hoạt động	Chuẩn đầu ra	Thành phần đánh giá
1 (4 tiết)	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 Cấu trúc dữ liệu (dữ liệu chi tiết hiện hành, chi tiết cũ, tổng hợp sơ bộ, mức cao, siêu dữ liệu) 1.5 Dòng dữ liệu trong kho dữ liệu 1.6 Kiến trúc kho dữ liệu (theo 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	G1.1 G2.1	Dạy: giải thích nhu cầu sử dụng dữ liệu trong doanh nghiệp, trình bày các 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, sử dụng công cụ tìm kiếm online và trả lời câu hỏi giảng viên. Học ở nhà: hình thành nhóm đồ án môn học và lập kế hoạch thực hiện.	A2
2, 3	Chương 2: 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	G2.1	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	A1.1 A2

	độ đo,... 2.3 Mô hình dữ liệu luận lý 2.4 Mô hình dữ liệu vật lý 2.5 Vấn đề hiệu suất		trả lời câu hỏi, làm bài tập trên lớp Học ở nhà: bài tập 1 + thảo luận nhóm đồ án môn học	
4, 5	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 3.5 Duy trì dữ liệu trong kho dữ liệu 3.6 Phát triển kho dữ liệu	G2.2	Dạy: trình bày phương pháp trích xuất dữ liệu, phương pháp biến đổi, nạp dữ liệu, minh họa bằng phần mềm SSIS, sự phát triển kho dữ liệu. Học ở nhà: bài tập 2, thực hiện đồ án MH.	A1.2
6	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	G2.2	Dạy: trình bày các pp mô phỏng csdl dạng khối, phân tích dữ liệu đa chiều, minh họa bằng phần mềm SSAS. Thực hiện đồ án.	A2
7, 8	4.2 Ngôn ngữ truy vấn đa chiều	G2.2	Dạy: trình bày ngôn ngữ truy vấn đa chiều, minh họa. Học ở lớp: làm bài tập trên lớp qua case study Học ở nhà: bài tập 3, thực hiện đồ án.	A1.3 A2
9	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.	G2.2	Dạy: triển khai xây dựng kho dữ liệu mẫu Adventure Works, đặt câu hỏi. Học ở lớp: sinh viên thảo luận, tự đánh giá lại bài tập của mình Thực hiện đồ án.	

10	Chương 5: HỖ TRỢ RA QUYẾT ĐỊNH (DSS) 5.1 Giới thiệu các phương pháp khai phá dữ liệu (data mining) 5.2 Ứng dụng công cụ BI hỗ trợ ra quyết định - Xác định kỹ thuật khai phá dữ liệu. - Cấu trúc khai phá (input, output, các tham số cho dữ liệu huấn luyện và dữ liệu thử,...)	G2.2	Dạy: Thuyết giảng, demo công cụ BI hỗ trợ ra quyết định, và đặt câu hỏi cho sinh viên. Học ở lớp: trả lời câu hỏi, thảo luận, tham gia xây dựng bài học. Thực hiện đồ án.	A2
11	Góp ý bài tập nhóm (đồ án) Ôn tập		Dạy: - Góp ý, nhận xét, trên các đồ án (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 30% số 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	Nội dung	CDRMH	Tỷ lệ %
A1. Bài tập về nhà 1,2,3,4 (A1.1, A1.2,	Triển khai xây dựng kho dữ liệu từ cơ sở dữ liệu mẫu Adventure Works và xử lý phân tích dữ liệu bằng công cụ BI và SQL Server DBMS.	G1.1, G2.1, G2.2	50%

A1.3) + ĐỒ ÁN MÔN HỌC	<ul style="list-style-type: none"> - 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 theo chủ đề tùy chọn. - 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). - A1.3: Xử lý phân tích dữ liệu trực tuyến từ kho, hỗ trợ ra quyết định (Quá trình SSAS và DSS). - Đồ án Môn học: Phân tích, thiết kế hoàn chỉnh một ứng dụng kho dữ liệu trong doanh nghiệp và hỗ trợ ra quyết định. 		
A2. Lý thuyết cuối kỳ	Tự luận	G2.1, G2.2	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.
- [3]. SQL Server 2019 Tutorials: Analysis Services - Multidimensional Modeling
- [4]. SQL Server 2019 Tutorials: Reporting Services (SSRS)
- [5]. SQL Server 2019 Tutorials: Analysis Services - Data Mining
- [6]. Google Data Studio Tutorials
- [7]. Power BI Tutorials

Tham khảo:

- [8]. E.G. Mallach, “Decision Support and Data Warehouse systems”, 2001.
- [9]. Marco Frailis, Data Management and Mining in Astrophysical Databases, 2003.
- [10]. Paulraj Ponniah, “Data Warehousing Fundamentals”, John Willey, 2003
- [11]. R. Kimball, “The Data Warehouse Toolkit”, John Willey, 2004.

9. PHẦN MỀM HAY CÔNG CỤ HỖ TRỢ THỰC HÀNH

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

Trưởng khoa/ bộ môn

Giảng viên