

BÀI 1: GIỚI THIỆU MÔN HỌC

KHO DỮ LIỆU VÀ OLAP

Giảng viên: ThS. Đỗ Thị Minh Phụng

Nội dung

- Mục tiêu môn học
- Nội dung môn học
- Phần mềm sử dụng
- Đánh giá kết quả môn học
- Tài liệu tham khảo

Mục tiêu môn học (chi tiết trong đề cương môn học) (1)

- Kết thúc môn học sinh viên nắm vững:
 - Kỹ năng phân tích nhu cầu báo cáo trực tuyến phục vụ doanh nghiệp.
 - Các khái niệm về kho dữ liệu, kiến trúc kho dữ liệu
 - Mô hình dữ liệu đa chiều (kỹ năng mô phỏng CSDL dạng khối), kỹ năng truy vấn dữ liệu đa chiều

Mục tiêu môn học (chi tiết trong đề cương môn học) (2)

- Biết cách sử dụng các công cụ BI của SQL Server:
 - Dịch vụ tích hợp dữ liệu SSIS
 - Dịch vụ phân tích dữ liệu SSAS
 - Dịch vụ báo biểu SSRS
 - Hỗ trợ ra quyết định (DSS, Data mining)
- Biết cách thiết kế một kho dữ liệu trong quản lý
- Sử dụng ngôn ngữ truy vấn đa chiều MDX
- Phân tích dữ liệu hỗ trợ ra quyết định

Nội dung môn học

- Tổng quan về kho dữ liệu
- Kiến trúc kho dữ liệu
- Phân tích, thiết kế kho dữ liệu (mô hình dữ liệu đa chiều, các phương pháp thiết kế, mức luận lý, vật lý)
- Tích hợp dữ liệu vào kho dữ liệu
- Phân tích dữ liệu đa chiều (Kỹ thuật OLAP)
- Ngôn ngữ MDX
- Phân tích dữ liệu hỗ trợ báo cáo theo nhu cầu doanh nghiệp và hỗ trợ ra quyết định

Đánh giá môn học

- Thi lý thuyết: cuối kỳ 50%
- Đồ án môn học và bài tập 50%
 - Theo nhóm: tối đa 2 sinh viên

Tài liệu tham khảo

- ▶ Bài giảng môn Kho dữ liệu
- ▶ W.H. Inmon, Building the Data Warehouses, Willey Dreamtech, 2004.
- ▶ R. Kimball, The Data Warehouse Toolkit, John Willey, 2004.
- ▶ Marco Frailis, Data Management and Mining in Astrophysical Databases, 2003.
- ▶ Paulraj Ponniah, Data Warehousing Fundamentals, John Willey, 2003
- ▶ E.G. Mallach, Decision Support and Data Warehouse systems, 2001.

Phần mềm hỗ trợ môn học

- ▶ Phần mềm yêu cầu:
 - SQL Server 2008 trở lên, cài đặt chức năng SQL Server Business Intelligence Development Studio được tích hợp vào trong Visual Studio
 - MS Excel
 - Cơ sở dữ liệu: NorthWind, AdventureWorks,...và các dataset từ các link sau:
- ☐ <http://snap.stanford.edu/data/index.html>
- ☐ http://catalogue.data.gov.bc.ca/dataset?download_audience=Public

Quy định đồ án môn học

- ▶ Nhóm tối đa 2 sinh viên.
- ▶ Hạn chót làm tiểu luận đến tuần lễ kết thúc môn.
- ▶ Hình thức nộp:
 - 01 báo cáo bản giấy kèm đĩa CD ghi nội dung tất cả kết quả làm việc của nhóm.

ĐÁNH GIÁ ĐỒ ÁN MÔN HỌC (chi tiết file đính kèm)

- Mô tả đề tài, kế hoạch thực hiện
- Khảo sát CSDL nguồn, nghiệp vụ của doanh nghiệp
- Thiết kế kiến trúc kho dữ liệu cho bài toán mà nhóm đề xuất, trích xuất, tích hợp và nạp dữ liệu vào kho từ các CSDL tác nghiệp (quá trình SSIS).
- Phân tích dữ liệu trực tuyến – OLAP (sử dụng công cụ Analysis Services của SQL Server và MS Excel) (SSAS).
- Phân tích dữ liệu sử dụng ngôn ngữ truy vấn đa chiều MDX.
- Sử dụng công cụ data mining trên kho dữ liệu để hỗ trợ ra quyết định.

Yêu cầu cụ thể

- 1. Mô tả chủ đề và nguồn dữ liệu gốc để làm đề tài.
- 2. Quá trình SSIS (3.0 điểm)
- 3. Quá trình SSAS: sử dụng ngôn ngữ MDX và thao tác tool (tối thiểu 15 truy vấn, bao gồm 4 loại truy vấn: roll up, drill down, slice and dice, pivot, các hàm hỗ trợ trong file hướng dẫn phần MDX) (3.0 điểm)
- 4. Dùng công cụ Pivot table của Excel thực hiện tương tự như quá trình SSAS (cùng câu truy vấn, thực hiện 5 câu) (1.0 điểm)

Yêu cầu cụ thể

- ▶ 5. Quá trình Data mining (thực hiện trên warehouse database, nếu kết quả không mining được tốt, chọn database khác) (2.0 điểm)
- ▶ 6. Report (3 report, có tối thiểu 1 report có group trên report) (1.0 điểm)
- ▶ 7. Phân công việc trong nhóm
- ▶ 8. Cấu trúc nộp đĩa:
 - a. Thư mục data: chứa dữ liệu gốc (nguồn)
 - b. Thư mục Source: project SSIS, SSAS, file Excel Pivot hướng dẫn sử dụng
 - c. Thư mục Database: chứa warehouse database (*.mdf, *.ldf)
 - d. Quay Video quá trình thực hiện
 - e. File Phân công công việc trong nhóm (Tên thành viên, công việc thực hiện)
 - f. Báo cáo đề tài (.docx)
