

## **CSC12107 – HTTT PV TRÍ TUỆ KINH DOANH**

### **ĐỒ ÁN THỰC HÀNH**

#### **1. Thông tin chung**

Mã số bài tập:	2025_DATH
Thời lượng dự kiến:	6 tuần
Deadline nộp bài:	10/12/2023
Hình thức:	Bài tập nhóm
Hình thức nộp bài:	Nộp qua Moodle môn học
GV phụ trách:	Hồ Thị Hoàng Vy, Tiết Gia Hồng, Nguyễn Ngọc Minh Châu
Thông tin liên lạc với GV:	<a href="mailto:hthvy@fit.hcmus.edu.vn">hthvy@fit.hcmus.edu.vn</a> , <a href="mailto:tghong@fit.hcmus.edu.vn">tghong@fit.hcmus.edu.vn</a> , <a href="mailto:nmmchau@fit.hcmus.edu.vn">nmmchau@fit.hcmus.edu.vn</a>

#### **2. Chuẩn đầu ra cần đạt**

Bài tập này nhằm mục tiêu đạt được các chuẩn đầu ra sau:

- G3.2: Giải thích vai trò và đặc trưng của kho dữ liệu, ý nghĩa của các bảng chiều và dữ kiện, các thuộc tính, phân cấp chiều trong lược đồ đa chiều
- G3.3: Thiết kế lược đồ chuẩn hoá, đa chiều (sao, bông tuyết) dựa vào dữ liệu hệ thống tác vụ và yêu cầu phân tích từ tình huống cho trước
- G4.2 Phân biệt được các ứng dụng khai phá dữ liệu vào kinh doanh thông minh. Giải thích đặc trưng và mục đích của khai phá dữ liệu, tầm quan trọng và một số kỹ thuật phổ biến
- G5.1 Triển khai quy trình ETL để rút trích dữ liệu từ nhiều nguồn, biến đổi, làm sạch dữ liệu, nạp dữ liệu vào KDL sử dụng công cụ SSIS
- G5.2 Xây dựng KDL đa chiều sử dụng công cụ SSAS và giải thích được lựa chọn phép toán OLAP phù hợp đối với 1 số yêu cầu phân tích.
- G5.3 Sử dụng một số công cụ biểu diễn dữ liệu (SSRS, powerBI, excel...) để biểu diễn kết quả phân tích, khai thác được (report, dashboard...)
- G5.4 Sử dụng SSAS và áp dụng các kỹ thuật mining tích hợp để thực hiện khai thác dữ liệu từ KDL xây dựng được.

### 3. Mô tả đề án

Đề tài tập trung xây dựng hệ thống kho dữ liệu (Data Warehouse) cho lĩnh vực hàng không, với dữ liệu từ các tập tin cung cấp (Airports, Airlines, Flights). Sinh viên thực hành đầy đủ các bước:

- Phân tích & thiết kế mô hình kho dữ liệu (schema).
- Nạp & cập nhật dữ liệu (ETL).
- Truy vấn dữ liệu & OLAP.
- Xây dựng Dashboard trực quan.
- Thực hiện Data Mining để khám phá tri thức.

#### 3.1 Yêu cầu:

- Dataset description: <https://www.kaggle.com/datasets/usdot/flight-delays?select=airlines.csv>
- Dataset: Moodle. **Lưu ý: Không sử dụng dataset trong link Kaggle**
- Truy vấn:
  - o Truy vấn tổng số chuyến bay theo tháng, quý, năm.
  - o Top 5 sân bay bận rộn nhất (có nhiều chuyến bay đi và đến nhất).
  - o Tỷ lệ chuyến bay đúng giờ (**On-Time Performance - OTP**)  $\pm 5$  theo sân bay.
  - o Tỷ lệ hủy chuyến theo nguyên nhân.
  - o Trung bình thời gian delay theo sân bay đi/đến
- Dashboard **gợi ý**:
  - o Quản lý:
    - Tổng số chuyến bay, OTP toàn ngành.
    - Tỷ lệ chuyến bị hủy, tỷ lệ delay > 15 phút.
    - Top 5 hãng hàng không theo OTP.
    - Top 5 sân bay có tỷ lệ delay cao nhất.
    - Xu hướng số chuyến bay theo tháng/quý.

Cấp 2: phân tích nguyên nhân:

- Nguyên nhân delay
- Thời gian delay trung bình theo nguyên nhân
- Xu hướng delay theo tháng/mùa
- Delay theo khung giờ bay
- Top sân bay đóng góp nhiều nhất vào delay/hủy
- o Sinh viên xác định yêu cầu phân tích & KPIs liên quan để tạo thêm dashboard quản lý tác nghiệp.
- o Lưu ý: dashboard cần xác định các filter, slicer phù hợp.
- Dự đoán khả năng chuyến bay bị delay.

#### 3.2 Cộng điểm (+0.5):

- Tìm hiểu và áp dụng AI để tự động diễn giải số liệu từ dashboard thành insight.
- Gợi ý: Power BI Smart Narrative

### 4. Các yêu cầu & quy định chi tiết cho bài nộp

- Kho dữ liệu hoàn chỉnh (báo cáo phân tích, thiết kế, script tạo csdl, project ETL dữ liệu)
- Script truy vấn OLAP, sql
- Dashboard (excel/powerBI)
- 1-2 Kỹ thuật mining áp dụng và báo cáo kết quả
- Báo cáo mô tả toàn bộ quá trình thiết kế, triển khai và phân tích.
- Kết luận các những hiểu biết giá trị, rút ra từ phân tích dữ liệu trên.

- Bài làm ghi rõ mã nhóm, mssv, họ tên vào đầu bài làm, phân công và đánh giá mỗi thành viên và toàn nhóm.
- Đặt tên: **MaNhom\_MSSV1\_MSSV2**

##### 5. Tài liệu tham khảo:

- Topic 01 – Hướng dẫn sử dụng SSIS
- Topic 02 – Hướng dẫn sử dụng SSAS - OLAP
- Topic 03 – Truy vấn OLAP với MDX
- PowerBI
  - <https://learn.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-smart-narrative>
  - <https://learn.microsoft.com/en-us/power-bi/create-reports/desktop-dimensional-model-report>

##### 6. Cách đánh giá

STT	Nội dung	Điểm
<b>1</b>	<b>Thiết kế &amp; cài đặt CSDL (NDS, DDS)</b>	<b>15%</b>
1.1	Thiết kế mô hình dữ liệu NDS đáp ứng yêu cầu lưu trữ của đồ án và được chuẩn hóa (tối thiểu 3NF)	5%
1.2	Thiết kế DDS có fact và dimension rõ ràng (có phân tích cụ thể từng yêu cầu và phân tích tổng hợp thiết kế nếu có)	8%
1.3	Cài đặt CSDL đúng thiết kế	2%
<b>2</b>	<b>Cài đặt ETL nạp dữ liệu</b>	<b>45%</b>
2.1	Xây dựng luồng Extract – Transform – Load đầy đủ, có log/ kiểm soát lỗi	15%
2.2	Thực hiện chuẩn hóa dữ liệu trước khi nạp	10%
2.3	Thực thi quy trình ETL không bị lỗi	10%
2.4	Tự động hóa ETL (schedule) có minh chứng	10%
<b>3</b>	<b>Report/OLAP/Dashboard</b>	<b>40%</b>
3.1	Tạo báo cáo tĩnh đúng yêu cầu và có hỗ trợ lọc theo tham số	10%
3.2	Thực hiện OLAP	10%
3.3	Xây dựng dashboard trực quan	15%
3.4	Hỗ trợ tương tác để khai thác trên dashboard	5%
<b>Tổng</b>		<b>100%</b>