

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN -  
ĐHQG-HCM**

**Khoa Công nghệ Thông tin**



**BÁO CÁO ĐỒ ÁN MÔN HỌC  
CSC12005 - HTTT PHỤC VỤ TRÍ TUỆ KINH DOANH**

**Giảng viên hướng dẫn**

ThS. Hồ Thị Hoàng Vy

ThS. Nguyễn Ngọc Minh Châu

ThS. Tiết Gia Hồng

**Thành viên nhóm** MSSV

Văn Bá Đức Kiên 22127218

Trần Quốc Phong 22127327

Lê Thị Thanh Thùy 22127411

Lê Ngọc Vũ 22127452

## Mục lục

1) Thông tin nhóm và phân công công việc .....	3
2) Thiết kế cơ sở dữ liệu .....	3
2.1) Thiết kế CSDL Nguồn (FLIGHTS_Source) .....	3
2.1.1) Các bảng trong Source .....	3
2.2) Metadata .....	3
2.3) Thiết kế CSDL Staging (FLIGHTS_Staging) .....	3
2.3.1) Các bảng trong Staging .....	3
3) ETL Pipeline .....	4
3.1) Source sang Stage .....	4
3.1.1) Tổng thể package .....	4
3.1.2) Execute SQL Task: Truncate Staging Tables .....	4
3.1.3) Execute SQL Task: Set CET .....	4
3.1.4) Execute SQL Task: Get Flights LSET, CET .....	4
3.1.5) Execute SQL Task: Set Flights LSET .....	4
3.1.6) Data Flow Task: Extract Airports, Airlines, Flights .....	5
Tài liệu tham khảo .....	6

## 1) Thông tin nhóm và phân công công việc

Mã nhóm	iSS-01			
Lớp	22HTTT1			
MSSV	Họ tên	Phân công	Đánh giá	Tỉ lệ đóng góp
22127128	Văn Bá Đức Kiên	Thực thi SSIS cho hệ thống	Hoàn thành đầy đủ và đúng hạn	25%
22127327	Trần Quốc Phong	Thiết kế CSDL Stage	Hoàn thành đầy đủ và đúng hạn	25%
22127411	Lê Thị Thanh Thùy	Thiết kế CSDL Source	Hoàn thành đầy đủ và đúng hạn	25%
22127452	Lê Ngọc Vĩ	Thiết kế CSDL Metadata	Hoàn thành đầy đủ và đúng hạn	25%

## 2) Thiết kế cơ sở dữ liệu

### 2.1) Thiết kế CSDL Nguồn (FLIGHTS\_Source)

- Lưu trữ dữ liệu ở dạng gần giống với file gốc
- Dễ dàng kiểm tra và đối chiếu khi có lỗi ETL

#### 2.1.1) Các bảng trong Source

Dữ liệu được nạp từ:

- airlines.csv
- filtered\_airpoirt.csv
- filtered\_flights\_1.csv
- filtered\_flights\_2.csv
- filtered\_flights\_3.csv

Bảng	Mô tả	Ghi chú
Airlines	Danh sách hãng hàng không	
Airports	Thông tin sân bay	
Flights_1, Flights_2, Flights_3	Dữ liệu lịch sử chuyến bay chia theo file	

### 2.2) Metadata

CSDL FLIGHTS\_Metadata được sử dụng để theo dõi trạng thái nạp dữ liệu.

Bảng data\_flow

Cột	Mô tả
TableName	Tên bảng ETL
LSET	Thời điểm ETL thành công gần nhất
CET	Thời điểm ETL hiện tại

### 2.3) Thiết kế CSDL Staging (FLIGHTS\_Stage)

- Sao chép dữ liệu từ Source sang Stage
- Dùng Incremental Load:
  - Chỉ nạp dữ liệu mới hoặc dữ liệu thay đổi
  - Dựa trên trường CREATED, MODIFIED trong các bảng Flights\_x ở Source

#### 2.3.1) Các bảng trong Staging

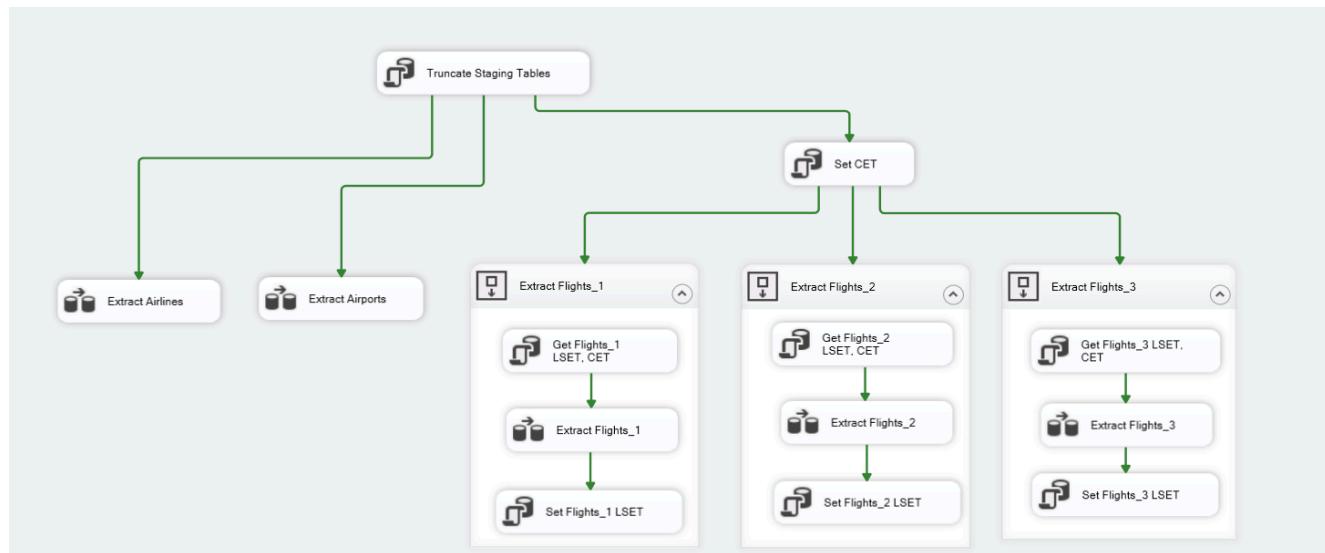
Bảng	Nguồn	Vai trò

STG_Airlines	Airlines	Lưu thông tin hãng hàng không
STG_Airports	Airports	Lưu thông tin sân bay
STG_Flights_1, STG_Flights_2, STG_Flights_3	Các bảng Flights tương ứng	Chuẩn bị hợp nhất và làm sạch dữ liệu

### 3) ETL Pipeline

#### 3.1) Source sang Stage

##### 3.1.1) Tổng thể package



##### 3.1.2) Execute SQL Task: Truncate Staging Tables

```

TRUNCATE TABLE STG_Airlines
TRUNCATE TABLE STG_Airports
TRUNCATE TABLE STG_Flights_1
TRUNCATE TABLE STG_Flights_2
TRUNCATE TABLE STG_Flights_3
  
```

##### 3.1.3) Execute SQL Task: Set CET

```

UPDATE data_flow SET CET = SYSDATETIME()
WHERE TableName IN (
    'Flights_1',
    'Flights_2',
    'Flights_3'
)
  
```

##### 3.1.4) Execute SQL Task: Get Flights LSET, CET

```

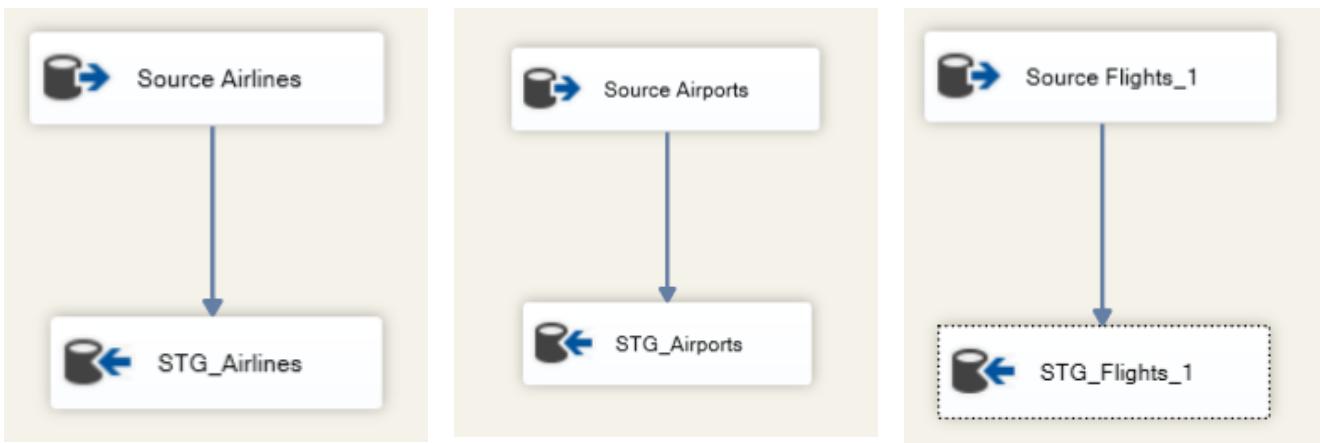
SELECT TOP 1 LSET, CET
FROM data_flow
WHERE TableName = 'Flights_1'
  
```

##### 3.1.5) Execute SQL Task: Set Flights LSET

```

UPDATE data_flow SET LSET = CET
WHERE TableName = 'Flights_1'
  
```

## 3.1.6) Data Flow Task: Extract Airports, Airlines, Flights



Đối với các bảng Flights\_x sẽ được áp dụng Incremental Load:

```
SELECT * FROM Flights_x
WHERE (CREATED >= LSET) AND (CREATED < CET)
OR (MODIFIED >= LSET) AND (MODIFIED < CET)
```

## Tài liệu tham khảo