

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN

----- & -----



BÁO CÁO SEMINAR

Tên Môn học: Hệ thống thông tin phục vụ trí tuệ kinh doanh

Mã Môn học: CSC12107

Giảng viên Hướng dẫn: Cô Hồ Thị Hoàng Vy

Sinh viên: 22127009 - Đinh Gia Hội Anh
22127162 - Phan Thành Quang Huy
22127163 - Trần Đan Huy
22127368 - Nguyễn Minh Sơn

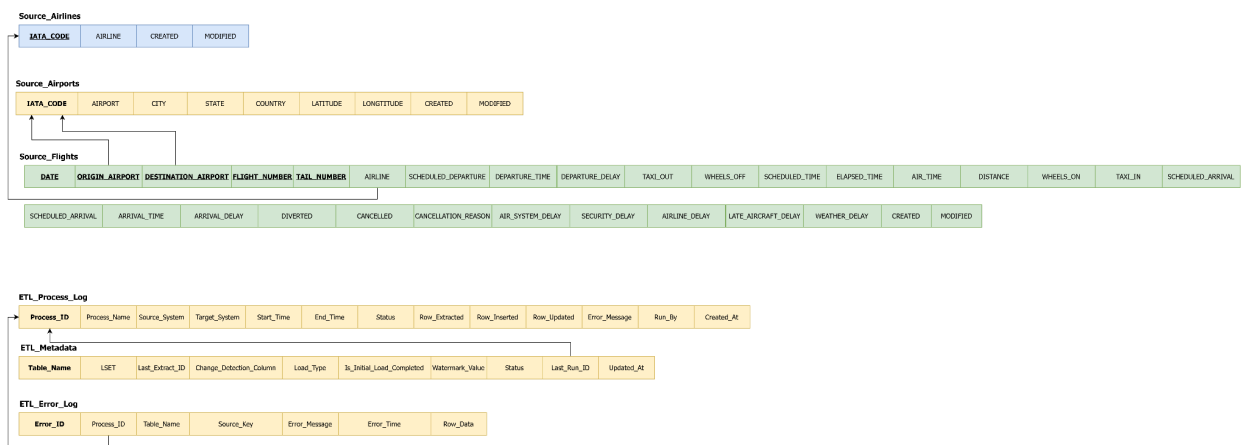
THÀNH PHỐ HỒ CHÍ MINH 2025 - 2026

MỤC LỤC

I. Data Modelling	3
1. Source	3
a) Source_Airlines	3
b) Source_Airports	3
c) Source_Flights	4
d) ETL_Process_Log	5
e) ETL_Metadata	5
f) ETL_Error_Log	6
2. Stage	6
3. NDS	7
a) NDS_Airlines	7
b) NDS_Airports	7
c) NDS_Flights	8
II. Quy trình ETL (Source → Stage)	9
1. Control Flow	10
2. Data Flow	11
3. Kết quả sau khi thực hiện source → stage	15

I. Data Modelling

1. Source



Hình trên là phân tích về data modelling của Source gồm các file csv. Ở giai đoạn này, ta sẽ giữ nguyên cấu trúc data thô của Source để đưa vào Stage.

a) Source_Airlines

Bảng này chứa thông tin danh mục về các hãng hàng không.

- **IATA_CODE**: Mã định danh 2 chữ cái của Hiệp hội Vận tải Hàng không Quốc tế (IATA) cho hãng bay (ví dụ: 'AA', 'VN'). Đây là Khóa Tự nhiên (Natural Key).
- **AIRLINE**: Tên đầy đủ của hãng hàng không (ví dụ: 'American Airlines', 'Vietnam Airlines').
- **CREATED**: Ngày bản ghi này được tạo trong hệ thống nguồn.
- **MODIFIED**: Ngày bản ghi này được cập nhật lần cuối trong hệ thống nguồn.

b) Source_Airports

Bảng này chứa thông tin danh mục về các sân bay.

- **IATA_CODE**: Mã định danh 3 chữ cái (IATA) cho sân bay (ví dụ: 'SGN', 'JFK'). Đây là Khóa Tự nhiên.

- **AIRPORT:** Tên đầy đủ của sân bay (ví dụ: 'Tan Son Nhat International Airport').
- **CITY:** Thành phố nơi sân bay tọa lạc.
- **STATE:** Bang (nếu áp dụng, ví dụ: 'CA' cho California).
- **COUNTRY:** Quốc gia (ví dụ: 'USA', 'Vietnam').
- **LATITUDE:** Vĩ độ (vị trí địa lý).
- **LONGITUDE:** Kinh độ (vị trí địa lý).
- **CREATED:** Ngày bản ghi này được tạo trong hệ thống nguồn.
- **MODIFIED:** Ngày bản ghi này được cập nhật lần cuối trong hệ thống nguồn.

c) Source_Flights

Bảng này chứa thông tin giao dịch chi tiết cho từng chuyến bay.

- **DATE:** Ngày diễn ra chuyến bay.
- **ORIGIN_AIRPORT:** Mã IATA của sân bay cất cánh.
- **DESTINATION_AIRPORT:** Mã IATA của sân bay hạ cánh.
- **FLIGHT_NUMBER:** Số hiệu chuyến bay (ví dụ: 'VN201').
- **TAIL_NUMBER:** Số hiệu đăng ký (số đuôi) của máy bay.
- **AIRLINE:** Mã IATA của hãng hàng không thực hiện chuyến bay.
- **SCHEDULED_DEPARTURE:** Giờ cất cánh dự kiến (theo lịch).
- **DEPARTURE_TIME:** Giờ cất cánh thực tế.
- **DEPARTURE_DELAY:** Thời gian trễ cất cánh (tính bằng phút, = DEPARTURE_TIME - SCHEDULED_DEPARTURE).
- **TAXI_OUT:** Thời gian máy bay lăn bánh từ cổng ra đường băng.
- **WHEELS_OFF:** Thời điểm máy bay chính thức cất bánh khỏi đường băng.
- **SCHEDULED_TIME:** Tổng thời gian bay dự kiến (theo lịch).
- **ELAPSED_TIME:** Tổng thời gian bay thực tế (từ lúc đóng cửa ở cổng đi đến mở cửa ở cổng đến).
- **AIR_TIME:** Thời gian bay thực tế trên không (từ WHEELS_OFF đến WHEELS_ON).
- **DISTANCE:** Khoảng cách của chuyến bay.
- **WHEELS_ON:** Thời điểm máy bay tiếp đất (chạm bánh xuống đường băng).
- **TAXI_IN:** Thời gian máy bay lăn bánh từ đường băng vào cổng.
- **SCHEDULED_ARRIVAL:** Giờ hạ cánh dự kiến (theo lịch).
- **ARRIVAL_TIME:** Giờ hạ cánh thực tế.
- **ARRIVAL_DELAY:** Thời gian trễ hạ cánh (tính bằng phút).
- **DIVERTED:** Cờ (flag) báo chuyến bay có bị chuyển hướng đến sân bay khác không (1 = Có, 0 = Không).
- **CANCELLED:** Cờ (flag) báo chuyến bay có bị hủy không (1 = Có, 0 = Không).
- **CANCELLATION_REASON:** Mã lý do hủy chuyến (nếu có).
- **AIR_SYSTEM_DELAY:** Thời gian trễ do lỗi hệ thống không lưu (ví dụ: thời tiết, kiểm soát không lưu).
- **SECURITY_DELAY:** Thời gian trễ do an ninh.
- **AIRLINE_DELAY:** Thời gian trễ do hãng hàng không (ví dụ: máy bay, phi hành đoàn).

- **LATE_AIRCRAFT_DELAY**: Thời gian trễ do máy bay đến muộn từ chuyến trước.
- **WEATHER_DELAY**: Thời gian trễ do thời tiết.
- **CREATED**: Ngày bản ghi này được tạo trong hệ thống nguồn.
- **MODIFIED**: Ngày bản ghi này được cập nhật lần cuối trong hệ thống nguồn.

Ngoài ra, để Stage thực hiện đúng cơ chế Incremental Load cũng như ghi nhận lại **events** những lúc chạy ETL để load data từ Source → Stage, ta cần các bảng hỗ trợ gọi là Metadata. Ở giai đoạn Source này, nhóm chúng em xin đưa ra thiết kế như sau

d) ETL_Process_Log

Bảng này ghi lại nhật ký mỗi lần chạy một quy trình ETL.

- **Process_ID**: ID duy nhất cho mỗi lần chạy.
- **Process_Name**: Tên quy trình (ví dụ: 'Load_NDS_Airlines').
- **Source_System**: Tên hệ thống nguồn.
- **Target_System**: Tên hệ thống đích.
- **Start_Time**: Thời điểm bắt đầu chạy.
- **End_Time**: Thời điểm kết thúc chạy.
- **Status**: Trạng thái ('Success', 'Failed', 'In Progress').
- **Row_Extracted**: Số dòng đã đọc từ nguồn.
- **Row_Inserted**: Số dòng đã chèn mới vào đích.
- **Row_Updated**: Số dòng đã cập nhật ở đích.
- **Error_Message**: Thông báo lỗi nếu quy trình 'Failed'.
- **Run_By**: Người dùng hoặc dịch vụ đã chạy quy trình.
- **Created_At**: Thời điểm hàng log này được tạo.

e) ETL_Metadata

Bảng này lưu "cấu hình" và "trạng thái" để ETL biết cần phải làm gì.

- **Table_Name**: Tên bảng đang được quản lý (ví dụ: 'Airlines', 'Flights1').
- **LSET (Last Successful Extract Time)**: Mốc thời gian quan trọng, lưu lại thời điểm tải thành công lần cuối.
- **Last_Extract_ID**: (Tương tự LSET) Mốc ID đã tải thành công lần cuối.
- **Change_Detection_Column**: Tên cột ở nguồn dùng để phát hiện thay đổi (ví dụ: 'ModifiedDate').
- **Load_Type**: Kiểu tải ('Incremental' - tăng dần, 'Full' - tải toàn bộ).
- **Is_Initial_Load_Completed**: Cờ (flag) báo đã tải lần đầu (toàn bộ lịch sử) hay chưa.
- **Watermark_Value**: Giá trị mốc (LSET hoặc Last_Extract_ID) đã lưu.
- **Status**: Trạng thái của lần tải cuối ('Success', 'Failed').
- **Last_Run_ID**: Con trỏ (khóa ngoại) trỏ đến Process_ID trong ETL_Process_Log của lần chạy cuối.

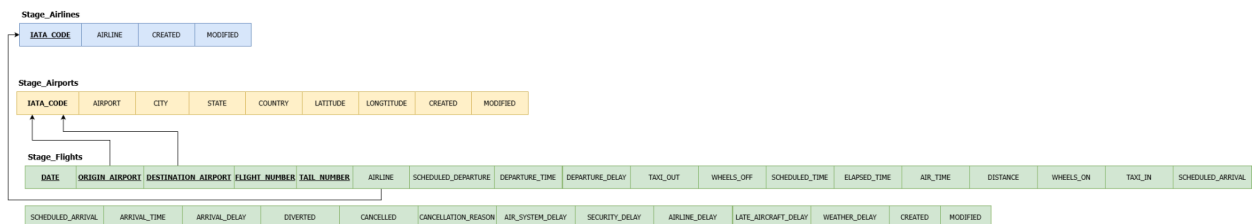
- **Updated_At**: Thời điểm dòng metadata này được cập nhật.

f) ETL_Error_Log

Bảng này dùng để "bắt" và ghi lại các lỗi ở row-level.

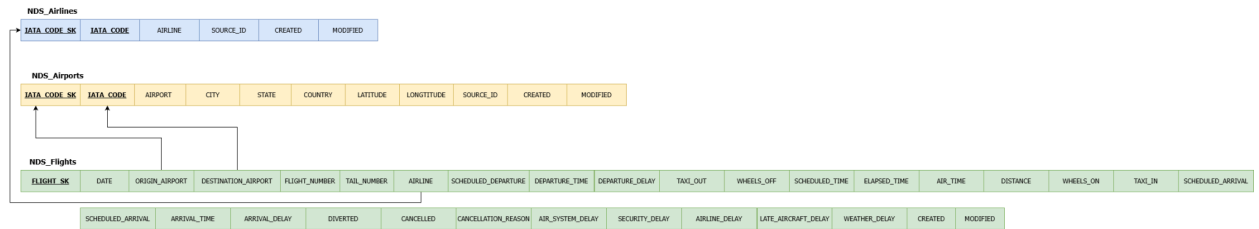
- **Error_ID**: ID duy nhất cho *mỗi hàng lỗi*.
- **Process_ID**: Quy trình (từ ETL_Process_Log) nào đã gây ra lỗi này.
- **Table_Name**: Bảng (nguồn/đích) chứa hàng lỗi.
- **Source_Key**: Khóa tự nhiên của hàng lỗi (ví dụ: 'SGN-HAN 2025-11-08 VN201') để truy vết.
- **Error_Message**: Lỗi chi tiết (ví dụ: 'Sai định dạng ngày', 'Không tìm thấy IATA_CODE').
- **Error_Time**: Thời điểm lỗi được ghi lại.
- **Row_Data**: Dữ liệu của *toàn bộ row bị lỗi* (thường ở dạng text/JSON) để kiểm tra và sửa thủ công.

2. Stage



- Cấu trúc 1:1 (Mirroring): Các bảng trong Stage bao gồm Stage_Airlines, Stage_Airports, và Stage_Flights. Chúng có cấu trúc cột y hệt các bảng tương ứng ở Giai đoạn Nguồn (Source).
- Mục tiêu Chính: Đóng vai trò là Landing Zone (Khu vực hạ cánh) – nơi chứa dữ liệu thô, sao chép 1:1 từ nguồn mà chưa áp dụng bất kỳ biến đổi nghiệp vụ nào.
- Tính Tạm thời: Dữ liệu trong Stage là dữ liệu tạm thời và thường được xóa sạch (Truncate) trước mỗi lần chạy ETL để đảm bảo rằng chỉ dữ liệu mới và sạch được nạp vào

3. NDS



Giai đoạn này, ta sẽ sinh ra Surrogate Key riêng cho NDS, để nó làm Primary Key cho toàn bộ data được đổ vào từ nhiều file source khác nhau. Vậy câu hỏi đặt ra là tại sao lại dùng SK thay vì NK.

- **Sự thay đổi của NK:** Nếu hãng bay thay đổi mã IATA (VN → VV), tất cả các bảng giao dịch (Flights) tham chiếu đến mã cũ sẽ bị sai lệch. Và SK là một mã số nguyên, duy nhất, không bao giờ thay đổi. Cụ thể là IATA_CODE_SK trong NDS_Airlines.
- **Phụ thuộc vào nguồn:** NK chỉ có ý nghĩa trong hệ thống nguồn đó. Nếu tích hợp dữ liệu từ hệ thống khác, có thể xảy ra trùng lặp NK. SK do Data Warehouse tự sinh ra, độc lập hoàn toàn với dữ liệu nguồn và đảm bảo tính duy nhất trên toàn hệ thống NDS. Ví dụ như FLIGHT_SK trong NDS_Flights.
- **Hiệu suất:** NK (như chuỗi ký tự dài, tổ hợp nhiều cột) làm chậm quá trình nối bảng (JOIN) trong cơ sở dữ liệu. SK là một số nguyên (integer), tối ưu hóa tốc độ JOIN và hiệu suất truy vấn trong quá trình phân tích dữ liệu

a) NDS_Airlines

- **IATA_CODE_SK:** (Cột mới) Khóa Thay thế (Surrogate Key). Đây là một mã ID số nguyên, duy nhất, do NDS tự sinh ra để làm khóa chính.
- **IATA_CODE:** Khóa Tự nhiên (Natural Key) từ nguồn, dùng để tra cứu và tham chiếu.
- **AIRLINE:** Tên hãng bay.
- **SOURCE_ID:** (Cột mới) Mã định danh hệ thống nguồn (cho biết bản ghi này đến từ đâu).
- **CREATED:** Ngày bản ghi này được chèn vào NDS.
- **MODIFIED:** Ngày bản ghi này được cập nhật lần cuối trong NDS.

b) NDS_Airports

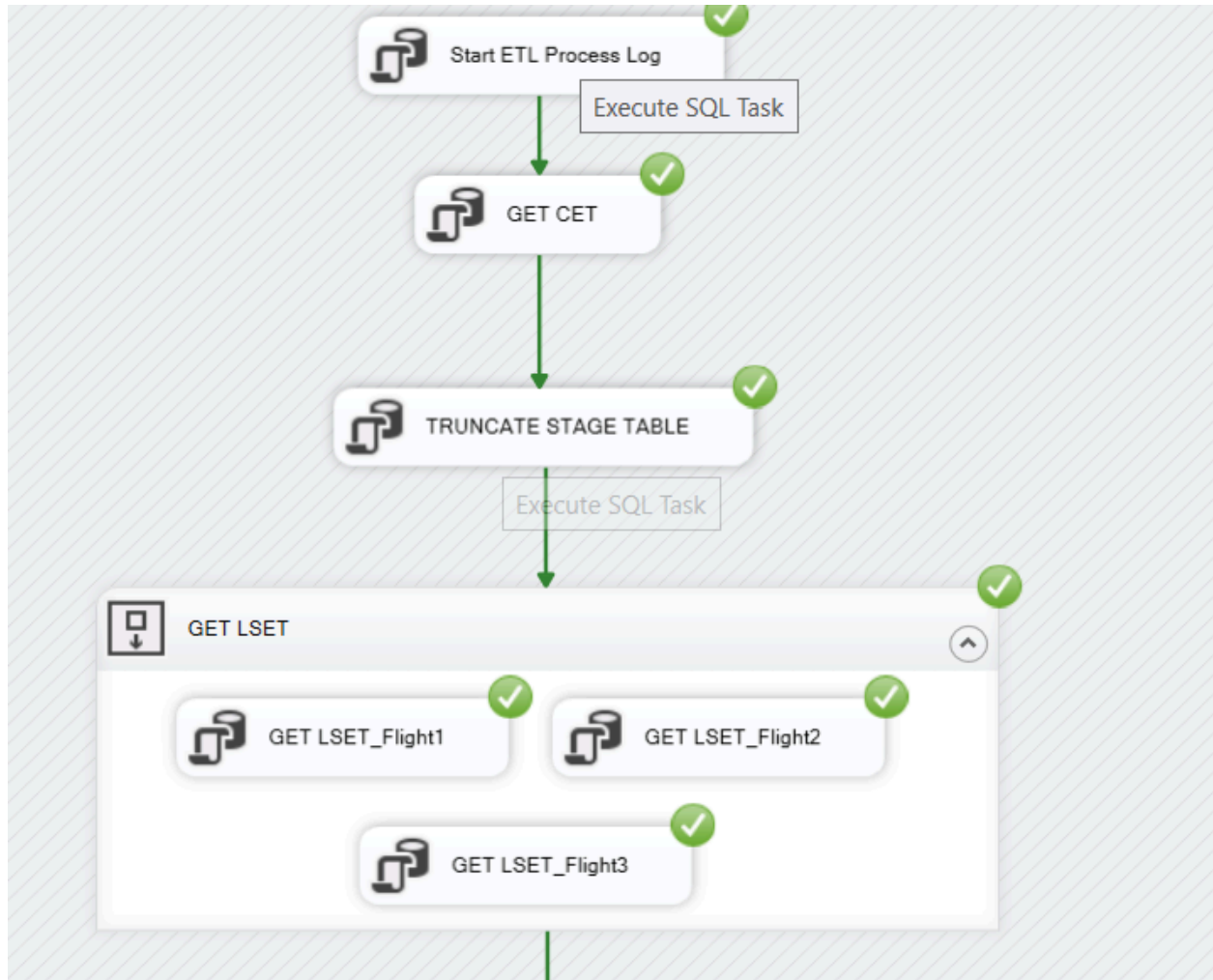
- **IATA_CODE_SK:** (Cột mới) Khóa Thay thế (Surrogate Key), khóa chính của bảng.
- **IATA_CODE:** Khóa Tự nhiên (Natural Key) từ nguồn.
- **AIRPORT:** Tên sân bay.
- **CITY:** Thành phố.
- **STATE:** Bang.

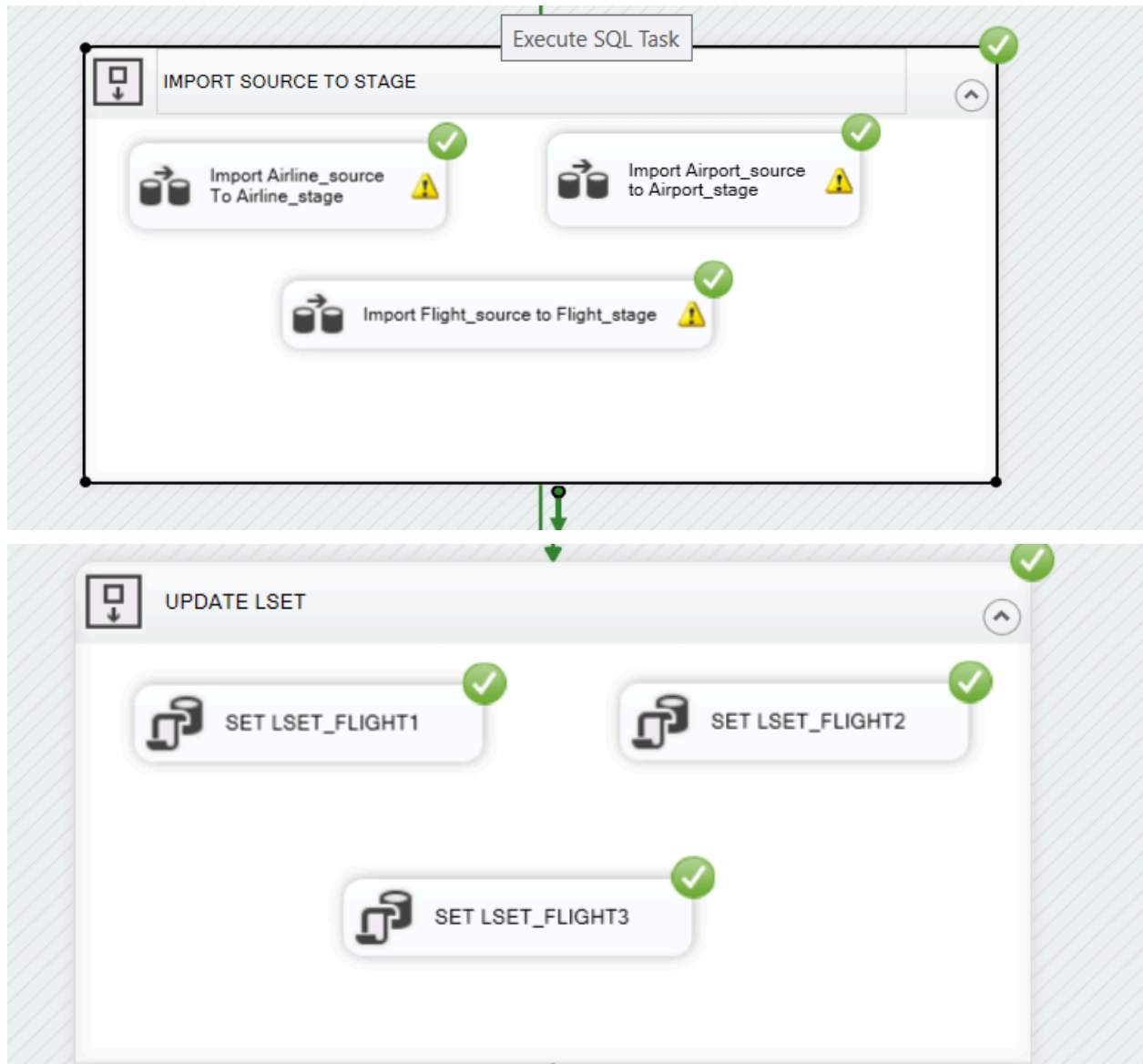
- **COUNTRY:** Quốc gia.
- **LATITUDE:** Vĩ độ.
- **LONGITUDE:** Kinh độ.
- **SOURCE_ID:** (Cột mới) Mã định danh hệ thống nguồn.
- **CREATED:** Ngày bản ghi này được chèn vào NDS.
- **MODIFIED:** Ngày bản ghi này được cập nhật lần cuối trong NDS.

c) NDS_Flights

- **FLIGHT_SK:** (Cột mới) Khóa Thay thế (Surrogate Key), khóa chính duy nhất của bảng sự kiện này.
- (Các cột nghiệp vụ): Tất cả các cột còn lại (từ DATE, ORIGIN_AIRPORT, DEPARTURE_DELAY, ARRIVAL_DELAY, DISTANCE, DIVERTED, CANCELLED... cho đến WEATHER_DELAY) được giữ nguyên từ nguồn/stage. Chúng mô tả chi tiết về sự kiện chuyến bay.

II. Quy trình ETL (Source → Stage)





1. Control Flow

Quy trình ETL từ Source vào Stage gồm các bước như sau:

1. Start ETL Process log

- Ghi nhận log khi bắt đầu chạy quá trình ETL.
- Các log này được dùng để keep track thông tin chi tiết mỗi lần quy trình extract và insert được bao nhiêu dữ liệu vào Stage.
- Các log này cũng cập nhập các lần ETL bị lỗi do về dữ liệu, hay về mặt kỹ thuật.

2. Set CET (Current Extract Time)

- Ghi nhận thời điểm extract hiện tại của phiên chạy.
- Dùng làm mốc thời gian để xác định dữ liệu mới phát sinh.

3. Truncate Tables

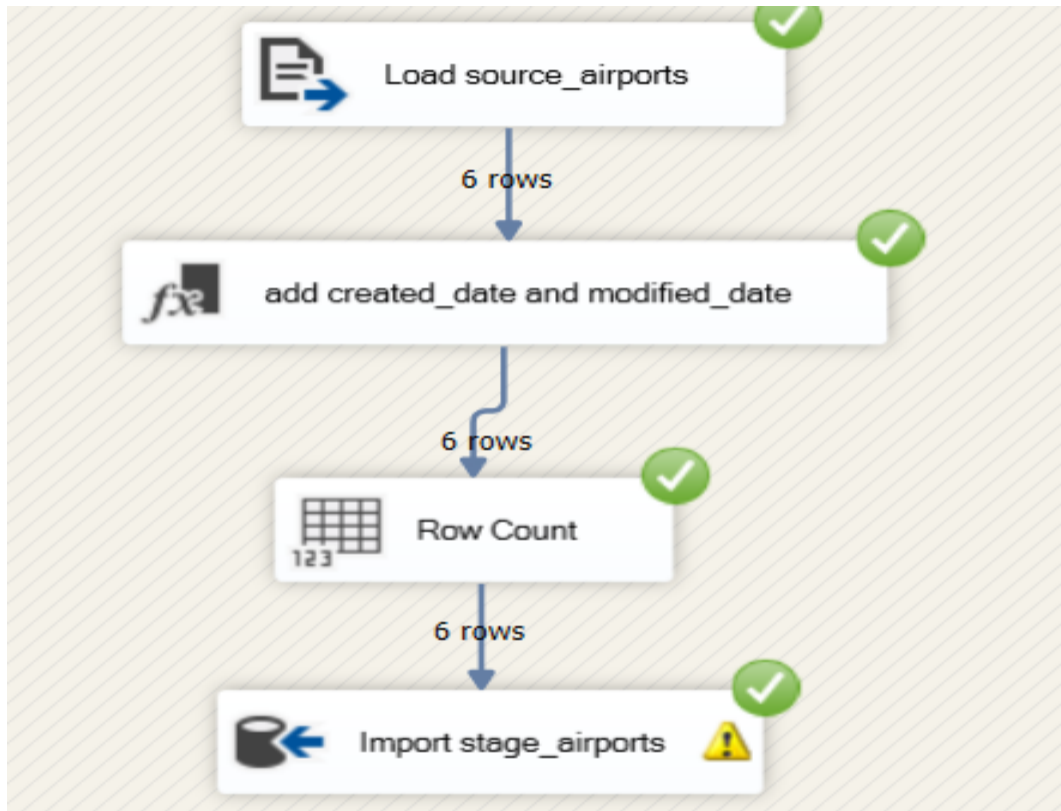
- Xóa toàn bộ dữ liệu tạm đang có trong các bảng Stage.

- Đảm bảo dữ liệu mới được nạp vào là **hoàn toàn sạch** (Zero-Duplication) và sẵn sàng cho các biến đổi tiếp theo mà không bị ảnh hưởng bởi dữ liệu từ các phiên chạy trước
- 4. Get LSET (Last Successful Extract Time)**
 - Hệ thống lấy lại thời điểm extract thành công gần nhất cho từng nguồn dữ liệu (được lưu trong ETL_Metadata):
 - Hãng bay (Airlines)
 - Sân bay (Airport)
 - Các nhóm chuyến bay (Flights 1, Flights 2, Flights 3)
 - Việc xác định LSET giúp hệ thống chỉ extract dữ liệu nằm trong khoảng thời gian từ LSET đến CET (Incremental Loading).
- 5. Trích dữ liệu từ Source sang Stage**
 - Tiến hành lấy dữ liệu mới phát sinh hoặc cập nhật từ các nguồn tương ứng:
 - Dữ liệu hãng bay
 - Dữ liệu sân bay
 - Dữ liệu chuyến bay
 - Dữ liệu sau khi lấy về sẽ được lưu tạm trong khu vực Stage để chuẩn bị cho các bước xử lý tiếp để đưa vào NDS.
- 6. Update LSET**
 - Sau khi extract hoàn tất, hệ thống cập nhật lại LSET trong ETL_Metadata bằng giá trị CET hiện tại.
 - Nhờ đó, ở những lần chạy tiếp theo chỉ cần xử lý dữ liệu mới phát sinh sau lần extract này.

2. Data Flow

Dưới đây là mô tả data flow của bước 4 - Trích dữ liệu từ Source sang Stage:

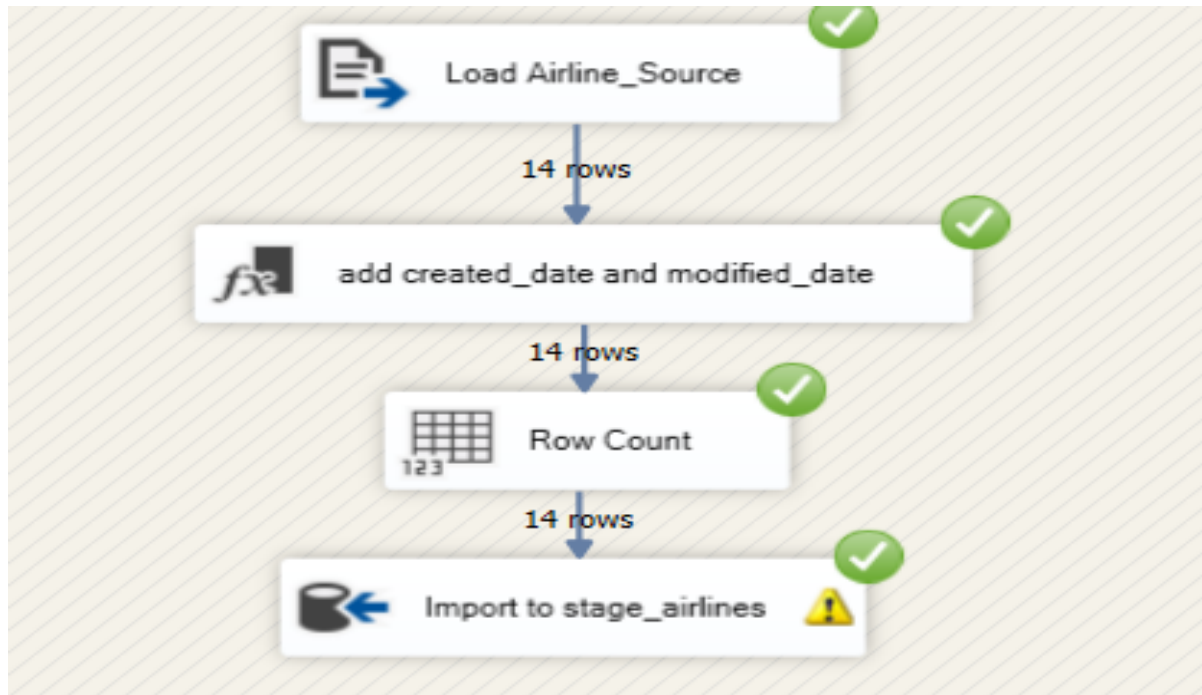
Extract Airport Source



Dữ liệu được lấy từ Airport Source (filtered_airpoint.csv), chứa thông tin về các sân bay. Sau khi extract, dữ liệu được nạp vào Stage_Airport.

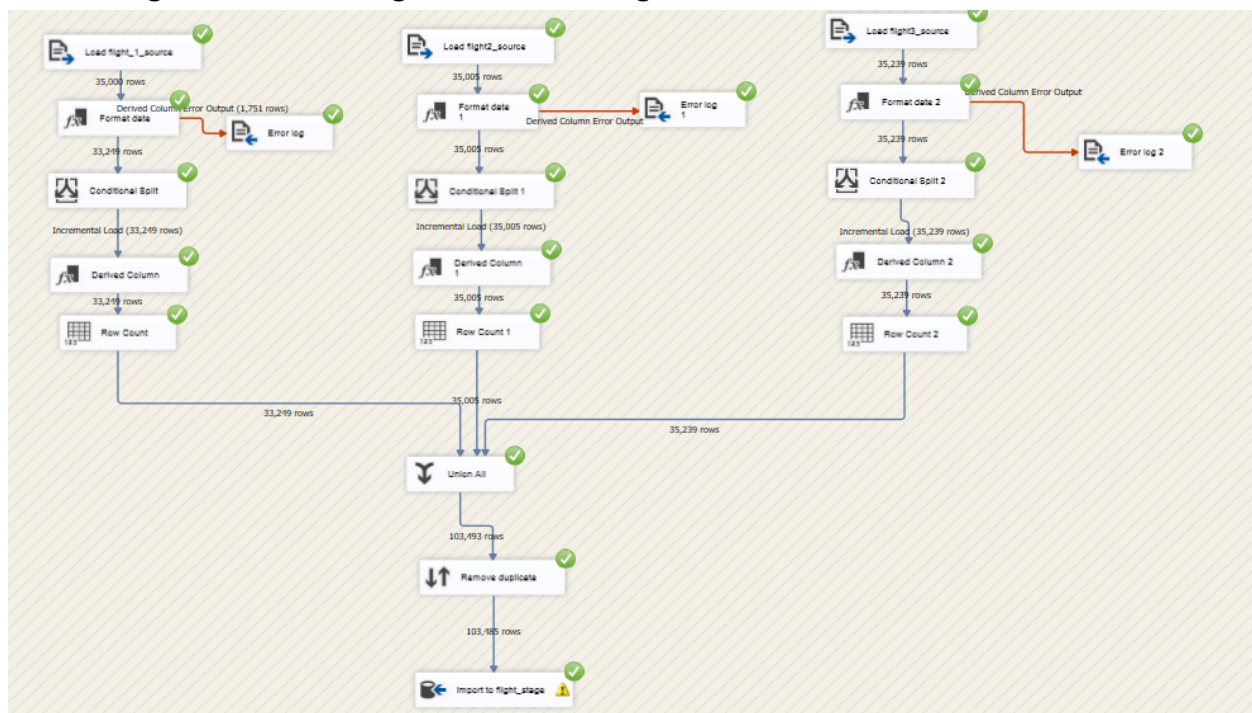
Các component được sử dụng:

- **Load source_airports:** Thực hiện hành động Extract cơ bản, đọc toàn bộ dữ liệu danh mục sân bay từ file nguồn (CSV). Bởi đây là bảng danh mục nhỏ (6 dòng trong lần chạy này), do đó việc tải toàn bộ (Full Load) là phương pháp tối ưu.
- **add created_date and modified_date:** Thêm các cột metadata CREATED và MODIFIED vào luồng dữ liệu. Để đảm bảo mọi bản ghi danh mục được ghi nhận thời điểm chính xác chúng được xử lý trong hệ thống ETL. Điều này quan trọng cho các cơ chế quản lý thay đổi (ví dụ: SCD Type 1/2) ở các tầng sau.
- **Row Count:** Đếm số lượng dòng dữ liệu đã qua xử lý (6 dòng). Để cập nhật số liệu vào ETL_Process_Log. Đây là bước kiểm toán cần thiết để đối soát: đảm bảo số dòng được đọc từ nguồn khớp với số dòng được chèn vào Stage, duy trì tính minh bạch của quy trình.
- **Import stage_airports:** Chèn dữ liệu đã làm sạch cơ bản vào bảng đích Stage_Airports. Để hoàn tất việc đưa dữ liệu thô 1:1 vào Stage, phục vụ vai trò Landing Zone.
- **Extract Airlines Source**



- Dữ liệu được lấy từ Airlines Source (airlines.csv), chứa thông tin về các hãng hàng không.
- Dữ liệu sau đó được lưu vào Stage_Airlines.
- Quá trình này sẽ thực hiện tương tự như ở airport

Extract Flights Source 1, Flights Source 2, Flights Source 3



Hệ thống trích xuất dữ liệu từ ba nguồn chuyến bay khác nhau (filtered_flights_1.csv, filtered_flights_2.csv, filtered_flights_3.csv).

Sau khi extract, dữ liệu được nạp vào bảng Stage_Flights.

Các component được sử dụng:

- **Load flight_1/2/3_source:** Trích xuất dữ liệu giao dịch từ ba file nguồn khác nhau. Để bắt đầu quá trình tích hợp dữ liệu chuyển bay phân tán thành một luồng xử lý đồng nhất.
- **Formal data / Derived Column Error Output (→ Error log 1/2):** Làm sạch Cấp độ Dòng (Row-level Cleansing). Để kiểm tra định dạng dữ liệu (ví dụ: ngày, số) và chuyển hướng các dòng không hợp lệ (ví dụ: ngày tháng sai, giá trị ngoài phạm vi) ra bảng ETL_Error_Log. Điều này giúp cách ly dữ liệu lỗi mà không làm dừng quy trình chính, đồng thời ghi lại để kiểm tra và sửa chữa thủ công sau này.
- **Conditional Split 1/2:** Phân tách luồng dữ liệu dựa trên các điều kiện nghiệp vụ. Có thể dùng để loại bỏ hoặc chuyển hướng các bản ghi không đủ điều kiện để phân tích (ví dụ: các bản ghi thiếu trường khóa quan trọng), giúp giảm tải dữ liệu ngay từ đầu.
- **Incremental Load:** Lọc dữ liệu theo điều kiện thời gian đã thiết lập (LSET < MODIFIED_DATE <= CET). Thực hiện cơ chế Tải Tăng Dần. Chỉ cho phép dữ liệu mới phát sinh hoặc cập nhật đi qua, đảm bảo hiệu suất và tránh xử lý lại dữ liệu lịch sử đã có.
- **Derived Column:** Thêm các cột tính toán hoặc metadata cần thiết. Thường dùng để thêm các cột metadata (như CREATED, MODIFIED) hoặc chuẩn bị Khóa Tự nhiên (NK) để dễ dàng xác định bản ghi trong bước loại bỏ trùng lặp.
- **Row Count 1/2/3:** Ghi nhận số lượng dòng đã qua các bước lọc và làm sạch (Incremental Load). Dùng để cập nhật số liệu vào ETL_Process_Log, thể hiện chính xác số lượng bản ghi tăng dần đã được trích xuất thành công từ mỗi nguồn.
- **Union All (103,493 rows):** Hợp nhất luồng dữ liệu từ ba nguồn (Flights 1, 2, 3) thành một luồng xử lý duy nhất. Để chuẩn bị cho bước loại bỏ trùng lặp và đảm bảo dữ liệu được xử lý đồng nhất trước khi nạp vào Stage.
- **Remove duplicate:** Loại bỏ các bản ghi trùng lặp. Do dữ liệu được hợp nhất từ nhiều nguồn và tải tăng dần, bước này là tối quan trọng để đảm bảo tính duy nhất của dữ liệu giao dịch trước khi nạp vào Stage.
- **Import to flight_stage:** Ghi dữ liệu sạch, đã hợp nhất và loại bỏ trùng lặp vào Stage_Flight. Hoàn tất việc đưa dữ liệu giao dịch đã qua tiền xử lý vào khu vực tạm

3. Kết quả sau khi thực hiện source → stage

Results Messages

	airline_code	airline_name	created	modified
1	AA	American Airlines Inc.	2025-11-09 09:12:28.4360000	2025-11-09 09:12:28.4360000
2	AS	Alaska Airlines Inc.	2025-11-09 09:12:28.4360000	2025-11-09 09:12:28.4360000
3	B6	JetBlue Airways	2025-11-09 09:12:28.4360000	2025-11-09 09:12:28.4360000
4	DL	Delta Air Lines Inc.	2025-11-09 09:12:28.4360000	2025-11-09 09:12:28.4360000
5	EV	Atlantic Southeast Airlines	2025-11-09 09:12:28.4360000	2025-11-09 09:12:28.4360000
6	F9	Frontier Airlines Inc.	2025-11-09 09:12:28.4360000	2025-11-09 09:12:28.4360000
7	HA	Hawaiian Airlines Inc.	2025-11-09 09:12:28.4360000	2025-11-09 09:12:28.4360000
8	MQ	American Eagle Airlines Inc.	2025-11-09 09:12:28.4360000	2025-11-09 09:12:28.4360000
9	NK	Spirit Air Lines	2025-11-09 09:12:28.4360000	2025-11-09 09:12:28.4360000
10	OO	Skywest Airlines Inc.	2025-11-09 09:12:28.4360000	2025-11-09 09:12:28.4360000
11	UA	United Air Lines Inc.	2025-11-09 09:12:28.4360000	2025-11-09 09:12:28.4360000
12	US	US Airways Inc.	2025-11-09 09:12:28.4360000	2025-11-09 09:12:28.4360000
13	VX	Virgin America	2025-11-09 09:12:28.4360000	2025-11-09 09:12:28.4360000
14	WN	Southwest Airlines Co.	2025-11-09 09:12:28.4360000	2025-11-09 09:12:28.4360000

Ảnh 3.1: Kết quả bảng Stage Airlines sau khi thực hiện source → stage

Results Messages

	airport_code	airport_name	city	state	country	latitude	logitude	created	modified
1	BWI	Baltimore-Washington International Airport	Baltimore	MD	USA	39.175400	-76.668200	2025-11-09 09:12:28.4580000	2025-11-09 09:12:28.4580000
2	CLT	Charlotte Douglas International Airport	Charlotte	NC	USA	35.214010	-80.943130	2025-11-09 09:12:28.4580000	2025-11-09 09:12:28.4580000
3	DFW	Dallas/Fort Worth International Airport	Dallas-Fort Worth	TX	USA	32.895950	-97.037200	2025-11-09 09:12:28.4580000	2025-11-09 09:12:28.4580000
4	EWB	Newark Liberty International Airport	Newark	NJ	USA	40.692500	-74.168660	2025-11-09 09:12:28.4580000	2025-11-09 09:12:28.4580000
5	LGA	LaGuardia Airport (Marine Air Terminal)	New York	NY	USA	40.777240	-73.872610	2025-11-09 09:12:28.4580000	2025-11-09 09:12:28.4580000
6	ORD	Chicago O'Hare International Airport	Chicago	IL	USA	41.979600	-87.904460	2025-11-09 09:12:28.4580000	2025-11-09 09:12:28.4580000

Ảnh 3.2: Kết quả bảng Stage Airports sau khi thực hiện source → stage

	FLIGHT_DATE	AIRLINE	FLIGHT_NUMBER	TAIL_NUMBER	ORIGIN_AIRPORT	DESTINATION_AIRPORT	SCHEDULED_DEPARTURE	DEPARTURE_TIME	DEPARTURE_DELAY	TAXI_OUT	WHEELS_OFF	SCHEDULED
1	2015-05-22 00:00:00.0000000	AA	1000	N3LCA	DFW	LGA	140	139	-1	33	212	203
2	2015-03-25 00:00:00.0000000	Click to select the whole column			BWI	ORD	944	943	-1	13	956	136
3	2015-06-23 00:00:00.0000000	UA	1001	N69804	ORD	DFW	608	615	7	18	633	147
4	2015-01-01 00:00:00.0000000	UA	1001	N23721	ORD	EWB	738	735	-3	12	747	125
5	2015-01-05 00:00:00.0000000	UA	1001	N37465	ORD	EWB	738	736	-2	11	747	125
6	2015-01-06 00:00:00.0000000	UA	1001	N39475	ORD	EWB	754	816	22	11	827	128
7	2015-01-07 00:00:00.0000000	UA	1001	N37255	ORD	EWB	600	708	68	14	722	125
8	2015-01-08 00:00:00.0000000	UA	1001	N37274	ORD	EWB	600	611	11	12	623	125
9	2015-01-09 00:00:00.0000000	UA	1001	N37281	ORD	EWB	754	814	20	15	829	128
10	2015-01-13 00:00:00.0000000	UA	1001	N38467	ORD	EWB	754	750	-4	12	802	128
11	2015-01-14 00:00:00.0000000	UA	1001	N68801	ORD	EWB	754	750	-4	12	802	128
12	2015-01-15 00:00:00.0000000	UA	1001	N73291	ORD	EWB	754	758	4	13	811	128
13	2015-01-16 00:00:00.0000000	UA	1001	N36280	ORD	EWB	754	749	-5	12	801	128
14	2015-01-20 00:00:00.0000000	UA	1001	N24202	ORD	EWB	754	750	-4	15	805	128
15	2015-01-21 00:00:00.0000000	UA	1001	N69834	ORD	EWB	754	750	-4	14	804	128
16	2015-01-22 00:00:00.0000000	UA	1001	N38268	ORD	EWB	754	750	-4	16	806	128

Query executed successfully. LAPTOP-9AKI0U5I (16.0 RTM) LAPTOP-9AKI0U5I\admin ... airline_stage 00:00:01 103,485 rows

Ảnh 3.3: Kết quả bảng Stage Flight sau khi thực hiện source → stage