

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO ĐỒ ÁN THỰC HÀNH

GD2: ETL NDS-> DDS, OLAP

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

23/11/2024 – 10/12/2024

MÃ HỌC PHẦN: CSC12107

Nhóm: 3

Thành viên:

ID	Họ tên
19127536	Võ Lâm Hải Quốc
21127211	Nguyễn Vũ Tường An
21127450	Võ Trung Tín
21127699	Lô Thủy Tiên

Giảng viên:

ThS. Hồ Thị Hoàng Vy

ThS. Tiết Gia Hồng

ThS. Nguyễn Ngọc Minh Châu

Báo cáo:

Lô Thủy Tiên

Thành phố Hồ Chí Minh – 2024

MỤC LỤC

THÔNG TIN VỀ ĐỒ ÁN.....	3
THÔNG TIN NHÓM.....	5
I. GIAI ĐOẠN : NDS-> DDS.....	6
1) Phân tích yêu cầu:.....	6
2) Cấu trúc bảng DDS:.....	18
a. Bảng DIM_States:.....	19
b. Bảng DIM_Counties:.....	20
c. Bảng DIM_Date:.....	20
d. Bảng DIM_Parameter:.....	21
e. Bảng FACT_AirQuality:.....	21
3) Data Flow:.....	22
II. PHÂN TÍCH QUY TRÌNH ETL:.....	23
1.1 Báo cáo giá trị AQI nhỏ nhất và lớn nhất cho từng bang trong mỗi quý của các năm.....	24
1.3. Báo cáo số ngày (number of days) và giá trị AQI trung bình (mean AQI) khi chất lượng không khí được đánh giá là "rất có hại" (very unhealthy) hoặc tồi tệ hơn cho từng bang (State) và từng quận (County).....	27
1.4 Đếm số ngày trong từng loại chất lượng không khí (Good, Moderate, Unhealthy, etc.) cho mỗi quận (County) thuộc bốn bang: Hawaii, Alaska, Illinois, và Delaware.....	28
1.6. Thiết kế một báo cáo để thể hiện xu hướng biến động của AQI trong năm cho bốn bang: Hawaii, Alabama, Utah, và Georgia.....	31
1.7. Báo cáo trung bình, độ lệch chuẩn, giá trị nhỏ nhất (min) và giá trị lớn nhất (max) của chỉ số AQI, nhóm theo Bang và Quận trong mỗi quý của năm. (bonus).....	32
PHÂN CÔNG CÔNG VIỆC.....	35
ĐÁNH GIÁ THÀNH VIÊN.....	39
TÀI LIỆU THAM KHẢO.....	40
<i>Công cụ và phần mềm hỗ trợ:.....</i>	<i>40</i>
<i>Tài liệu tham khảo:.....</i>	<i>41</i>

THÔNG TIN VỀ ĐỒ ÁN

Mã học phần: CSC12107

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

Tên : Đồ án thực hành - 2425.BI.DATH

Hình thức:

- Báo cáo (.doc, ppt)
- Source demo
- Video demo

Mô tả:

Dữ liệu sẽ được phân tích để phát hiện các xu hướng và mẫu trong chất lượng không khí của Hoa Kỳ từ năm 2021 đến 2023. Đồ án yêu cầu xây dựng một kho dữ liệu (DW) từ các nguồn dữ liệu thô, sau đó triển khai các quy trình ETL, thiết kế mô hình OLAP, thực hiện phân tích dữ liệu và tạo các báo cáo.

- **Dữ liệu:** Dữ liệu chất lượng không khí hàng ngày của EPA, phân chia theo quận từ năm 2021 đến 2023, kết hợp với dữ liệu địa lý và định nghĩa phân loại AQI.
- **Thiết kế kho dữ liệu:** Xây dựng các bảng chiều như Địa lý (State > County), Thời gian (Year > Quarter > Month > Day) và các chiều khác để phục vụ yêu cầu báo cáo.
- **Yêu cầu báo cáo và phân tích:**
 - Tạo các báo cáo biểu đồ để trình bày sự biến động AQI theo thời gian.
 - Phân tích các câu hỏi mở để đưa ra đánh giá về biến động AQI, ngày ô nhiễm nặng, và các yếu tố có ảnh hưởng đến chất lượng không khí.
 - Tích hợp bản đồ khu vực để biểu diễn trung bình AQI qua các khu vực.

- **Khai phá dữ liệu:** Ứng dụng các mô hình khai phá dữ liệu nhằm dự đoán chất lượng không khí trong các kỳ tới (Q1-2024, tháng 01-2024, ...), giải thích thuật toán, lý do chọn lựa, và trình bày kết quả.
- **Tổng quan và kết quả:** Đưa ra đánh giá tổng quan về chất lượng không khí ở các quận Hoa Kỳ vào năm 2023. Kết luận các thành tựu đạt được trong đồ án, cũng như đề xuất hướng cải thiện cho chất lượng không khí trong tương lai.

Mục tiêu:

- Thiết kế mô hình dữ liệu: Sử dụng mô hình ngôi sao hoặc bông tuyết để đáp ứng yêu cầu phân tích.
- Quy trình ETL: Sử dụng công cụ SSIS để thu thập, làm sạch và tích hợp dữ liệu từ các nguồn khác nhau vào kho dữ liệu.
- OLAP và Trục quan hóa: Sử dụng công cụ SSAS để khai thác các công nghệ OLAP cơ bản và tạo báo cáo bằng SSRS hoặc Excel.
- Khai phá dữ liệu: Áp dụng các thuật toán khai phá dữ liệu với SSAS để phân tích và dự đoán chất lượng không khí.

Giảng viên phụ trách: Cô Hồ Thị Hoàng Vy, Cô Tiết Gia Hồng, Cô Nguyễn Ngọc Minh Châu

THÔNG TIN NHÓM

Nhóm: 3

MSSV	Họ tên	Email	Ghi chú
19127536	Võ Lâm Hải Quốc	vlhquoc19@clc.fitus.edu.vn	
21127211	Nguyễn Vũ Tường An	nvtan21@clc.fitus.edu.vn	
21127450	Võ Trung Tín	vttin21@clc.fitus.edu.vn	
21127699	Lô Thủy Tiên	littien21@clc.fitus.edu.vn	

I. GIAI ĐOẠN : NDS-> DDS

1) Phân tích yêu cầu:

1.1. Báo cáo giá trị AQI nhỏ nhất và lớn nhất cho từng bang trong mỗi quý của các năm.

Gợi ý phân tích:

- Giá trị AQI biến động như thế nào trong suốt năm?
- Chú ý đến các giá trị lớn nhất và nhỏ nhất (max, min).
- Có giá trị nào bất thường, quá lớn hoặc quá nhỏ không?

a. Thông tin cần phân tích:

Báo cáo giá trị AQI (Air Quality Index) nhỏ nhất và lớn nhất

Phân tích theo từng bang (state)

Phân tích theo quý trong các năm từ 2021-2023

Chú ý đến sự biến động của giá trị AQI: Nhìn vào các giá trị cực đại và cực tiểu để nhận diện những thay đổi đột ngột hoặc bất thường (ví dụ: AQI quá cao hoặc quá thấp).

b. Bối cảnh sự kiện -> dimension table:

- Ai: Không có chiều người dùng trong báo cáo này
- Ở đâu: Dimension Geography (State, County)
- Cái gì: Dimension AQI_Category (phân loại chất lượng không khí)
- Khi nào: Dimension Time (Year, Quarter, Month, Day)

c. Đo lường (dữ kiện): Số lượng-> Fact Table

- Giá trị AQI hàng ngày từ bảng **NDS_AirData**.
- Giá trị Min AQI theo quý
- Giá trị Max AQI theo quý

d. Giá trị cần tính toán:

MIN(AQI) theo state và quarter

MAX(AQI) theo state và quarter

e. Phân cấp Chiều và Độ Mịn của Dữ Liệu:

- **Độ mịn dữ liệu (grain):** Mỗi dòng trong bảng sự kiện sẽ tương ứng với một giá trị AQI của một bang trong một quý
- **Phân cấp chiều:**
Time: Year > Quarter > Month > Day
Geography: State > County

1.2. Báo cáo giá trị trung bình (mean) và độ lệch chuẩn (standard deviation) của chỉ số AQI cho từng bang trong mỗi quý của các năm.

Gợi ý phân tích:

- Giá trị AQI biến động như thế nào trong suốt năm?
- Chú ý đến các giá trị: trung bình (mean), độ lệch chuẩn (std), lớn nhất (max) và nhỏ nhất (min).
- Có giá trị nào bất thường, quá lớn hoặc quá nhỏ không?

a. Thông tin cần phân tích:

- Chỉ số chất lượng không khí (AQI) của từng bang
- Phân tích theo quý và năm

b. Bối cảnh sự kiện -> dimension table:

- Ai : Cơ quan môi trường
- Ở đâu: Các bang tại Hoa Kỳ
- Cái gì: Chỉ số chất lượng không khí (AQI)
- Khi nào: Tại từng quý và năm

c. Đo lường (dữ kiện): Số lượng-> Fact Table

- Các chỉ số AQI:
 - Giá trị trung bình (Mean)
 - Độ lệch chuẩn (Standard Deviation)

d. Giá trị cần tính toán:

- Giá trị trung bình AQI theo bang và quý
- So sánh và phân tích biến động

e. Phân cấp Chiều và Độ Mịn của Dữ Liệu:

- Độ mịn dữ liệu (grain): Mỗi dòng trong bảng sự kiện sẽ tương ứng với một bản ghi AQI của một bang trong một quý
- Phân cấp chiều:
 - Không gian: Bang
 - Thời gian: Năm, Quý
 - Chỉ số: AQI (Mean, Std)

1.3. Báo cáo số ngày (number of days) và giá trị AQI trung bình (mean AQI) khi chất lượng không khí được đánh giá là "rất có hại" (very unhealthy) hoặc tồi tệ hơn cho từng bang (State) và từng quận (County).

Gợi ý phân tích:

- Tìm hiểu giới hạn AQI từ mức nào trở lên được coi là "rất có hại" hoặc tồi tệ hơn.
- Tính số ngày và giá trị trung bình AQI trong các khoảng đó cho mỗi bang và quận.

a. Thông tin cần phân tích:

Báo cáo số ngày (number of days) và giá trị AQI trung bình (mean AQI) khi chất lượng không khí được đánh giá là "rất có hại" (very unhealthy) hoặc tồi tệ hơn cho từng bang (State) và từng quận (County).

b. Bối cảnh sự kiện -> dimension table:

- Ai : Không áp dụng
- Ở đâu: Các bang và quận của Hoa Kỳ (state_name, county_name, state_code, county_code).
- Cái gì: Loại chất lượng không khí, đặc biệt tập trung vào các mức "Very Unhealthy" hoặc tồi tệ hơn.
- Khi nào: Phân tích dữ liệu theo ngày, tháng, năm (date_id, full_date, year, month, day)

c. Đo lường (dữ kiện): Số lượng-> Fact Table

- Fact_AirQuality: Chứa số ngày và giá trị trung bình AQI cho từng Bang và Quận.

d. Giá trị cần tính toán:

- Phân loại: Tìm các dòng dữ liệu có Category là "Very Unhealthy" hoặc "Hazardous".
- Số ngày (Number of days): Đếm số lượng ngày xuất hiện AQI thuộc các loại này.
- AQI trung bình (Mean AQI): Tính trung bình AQI cho mỗi Bang và Quận trong khoảng "Very Unhealthy" hoặc tồi tệ hơn.

e. Phân cấp Chiều và Độ Mịn của Dữ Liệu:

- **Độ mịn dữ liệu (grain):** Mỗi dòng trong Fact Table sẽ tương ứng với một lần đo AQI tại một quận vào một ngày cụ thể khi chất lượng không khí ở mức "Very Unhealthy" hoặc "Hazardous".
- **Phân cấp chiều:**
 - Thời gian: Năm → Tháng → Ngày.
 - Địa điểm: Bang → Quận.
 - Loại AQI: "Very Unhealthy" hoặc "Hazardous".

1.4. Đếm số ngày trong từng loại chất lượng không khí (Good, Moderate, Unhealthy, etc.) cho mỗi quận (County) thuộc bốn bang: Hawaii, Alaska, Illinois, và Delaware. (TIỀN)

Gợi ý phân tích:

- So sánh dữ liệu giữa các bang và quận.
- Tập trung vào phân bố các điều kiện không khí có hại (Unhealthy, Very Unhealthy, Hazardous).

- Kết luận: Xem xét sự khác biệt về chất lượng không khí giữa các bang/quận. Điều gì có thể gây ra sự khác biệt này? (Ví dụ: điều kiện địa lý, khí hậu, nguồn ô nhiễm).

a. Thông tin cần phân tích:

Đếm số ngày trong từng loại chất lượng không khí (Good, Moderate, Unhealthy, etc.) cho mỗi quận (County) thuộc bốn bang: Hawaii, Alaska, Illinois, và Delaware

b. Bối cảnh sự kiện -> dimension table:

- Ai : Không áp dụng
- Ở đâu: Các bang và quận của Hoa Kỳ (state_name, county_name, state_code, county_code).
- Cái gì: Phân loại chất lượng không khí (category_id, category_name, color_code, description)
- Khi nào: Phân tích dữ liệu theo ngày, tháng, năm (date_id, full_date, year, month, day)

c. Đo lường (dữ kiện): Số lượng-> Fact Table

- Số ngày trong từng loại chất lượng không khí (Good, Moderate, Unhealthy, etc.) cho mỗi quận (County) thuộc bốn bang

d. Giá trị cần tính toán:

Sử dụng hàm COUNT hoặc SUM trong quá trình ETL để tính số lượng ngày ở mỗi mức AQI trong mỗi quận và bang.

e. Phân cấp Chiều và Độ Mịn của Dữ Liệu:

- Độ mịn dữ liệu (grain): Mỗi dòng trong bảng sự kiện AQI_Fact tương ứng với số ngày ở mỗi mức chất lượng không khí cho từng quận vào một ngày cụ thể.
- Phân cấp chiều:
 - Location: State → County.
 - Time: Year → Month → Day.
 - AQI Category: Category Name.

1.5. Tính giá trị trung bình AQI theo từng quý cho bốn bang: Hawaii, Alaska, Illinois, và Delaware. (QUỐC)

Gợi ý phân tích:

- So sánh dữ liệu AQI trung bình giữa các bang qua các quý trong năm.
- Chú ý đến biến động trong từng bang và giữa các bang.
- Kết luận: Đưa ra nhận xét về xu hướng biến động AQI theo mùa hoặc theo điều kiện khí hậu của từng bang (ví dụ: mùa đông ở Alaska có thể ảnh hưởng khác so với mùa hè ở Hawaii).

a. Thông tin cần phân tích:

Giá trị trung bình AQI theo quý

Chỉ tập trung vào 4 bang: Hawaii, Alaska, Illinois, và Delaware

Phân tích xu hướng theo mùa và điều kiện khí hậu

b. Bối cảnh sự kiện -> dimension table:

- Ai: Không có chiều người dùng trong báo cáo này
- Ở đâu: Dimension Geography (State) với filter cho 4 bang cụ thể
- Cái gì: Dimension AQI_Category (phân loại mức độ chất lượng không khí)
- Khi nào: Dimension Time (Year, Quarter)

c. Đo lường (dữ kiện): Số lượng-> Fact Table

- Giá trị AQI hàng ngày
- Giá trị trung bình AQI theo quý (AVG_AQI)

d. Giá trị cần tính toán:

AVG(AQI) theo state và quarter

So sánh giữa các bang

Phân tích xu hướng theo mùa

e. Phân cấp Chiều và Độ Mịn của Dữ Liệu:

- **Độ mịn dữ liệu (grain):** Mỗi dòng trong bảng sự kiện sẽ tương ứng với giá trị trung bình AQI của một bang trong một quý
- **Phân cấp chiều:** Time: Year > Quarter > Month > Day

- **Geography: State (filtered to 4 states only)**

1.6. Thiết kế một báo cáo để thể hiện xu hướng biến động của AQI trong năm cho bốn bang: Hawaii, Alaska, Illinois, và California. (TÍN)

Gợi ý phân tích:

- Đưa ra nhận xét về **xu hướng biến động** của chỉ số AQI.
- Cân nhắc các yếu tố có thể ảnh hưởng đến sự biến động này, như **thời tiết, khí hậu, mùa** hoặc **nguồn ô nhiễm** đặc thù của từng bang.

a. Thông tin cần phân tích:

Báo cáo để thể hiện xu hướng biến động của AQI trong năm cho bốn bang: Hawaii, Alaska, Illinois, và California.

b. Bối cảnh sự kiện -> dimension table:

- Ai : Không áp dụng
- Ở đâu: Các bang của Hoa Kỳ (state_name, state_code).
- Cái gì: Chứa thông tin về chỉ số AQI và loại chất lượng không khí.
- Khi nào: Ngày, tháng, và năm đo AQI.

c. Đo lường (dữ kiện): Số lượng-> Fact Table

- Fact_AirQuality: Lưu trữ dữ liệu về giá trị AQI hàng ngày để theo dõi xu hướng trong năm.

d. Giá trị cần tính toán:

- AQI trung bình theo tháng: Tính trung bình chỉ số AQI theo từng tháng trong năm.
- Xu hướng biến động: Dựa vào các giá trị trung bình theo tháng để xác định xu hướng (tăng, giảm, hoặc ổn định).

e. Phân cấp Chiều và Độ Mịn của Dữ Liệu:

- **Độ mịn dữ liệu (grain):** Mỗi dòng trong Fact Table sẽ tương ứng với một lần đo AQI tại một bang vào một ngày cụ thể.
- **Phân cấp chiều:**
 - **Thời gian:** Năm → Tháng → Ngày.

- **Địa điểm:** Bang.
- **Chỉ số AQI:** Giá trị hàng ngày.

1.7. Báo cáo trung bình, độ lệch chuẩn, giá trị nhỏ nhất (min) và giá trị lớn nhất (max) của chỉ số AQI, nhóm theo Bang và Quận trong mỗi quý của năm. (bonus)

Gợi ý phân tích:

- Chú ý đến các giá trị: trung bình (mean), độ lệch chuẩn (std), lớn nhất (max), nhỏ nhất (min).
- Có giá trị nào bất thường (quá lớn hoặc quá nhỏ) không?
- So sánh độ lệch chuẩn giữa câu hỏi 1 và câu hỏi 2 và giải thích sự khác biệt.

a. Thông tin cần phân tích:

- Chỉ số chất lượng không khí (AQI)
- Phân tích chi tiết theo từng bang (State) và quận (County)
- Thống kê theo từng quý và từng năm

b. Bối cảnh sự kiện -> dimension table:

- Ai: Các trạm đo chất lượng không khí
- Ở đâu: Các quận thuộc các bang tại Hoa Kỳ
- Cái gì: Chỉ số chất lượng không khí (AQI)
- Khi nào: tại từng quý qua từng năm

c. Đo lường (dữ kiện): Số lượng-> Fact Table

- Các chỉ số thống kê AQI:
 - Giá trị trung bình (Mean)
 - Độ lệch chuẩn (Standard Deviation)
 - Giá trị lớn nhất (Max)
 - Giá trị nhỏ nhất (Min)

d. Giá trị cần tính toán:

- Trung bình AQI theo từng quận và bang

- Độ lệch chuẩn AQI
- So sánh và phân tích biến động
- Xác định các giá trị bất thường

e. Phân cấp Chiều và Độ Mịn của Dữ Liệu:

- **Độ mịn dữ liệu (grain):** Mỗi dòng trong bảng sự kiện sẽ tương ứng với một bản ghi AQI của một quận trong một quý
- **Phân cấp chiều:**
 - Không gian: Bang, Quận
 - Thời gian: Năm, Quý
 - Chỉ số: AQI (Mean, Std, Max, Min)

1.8. Tạo một thuộc tính mới, DayLightSaving, trong bảng phù hợp.

DayLightSaving có hai giá trị: (bonus) (TIỀN)

- True: Từ ngày 12 tháng 3, 2023 đến ngày 5 tháng 11, 2023
- False: Trong khoảng thời gian còn lại.

Báo cáo **giá trị AQI trung bình** theo **Bang**, **Loại chất lượng không khí (Category)**, **DayLightSaving** qua các năm.

Gợi ý phân tích:

- Có sự khác biệt đáng chú ý nào về **chất lượng không khí** trong **khoảng thời gian tiết kiệm ánh sáng ban ngày (Daylight Saving)** so với những thời gian khác không?
- So sánh chất lượng không khí giữa các khoảng thời gian **Daylight Saving** và **non-Daylight Saving**.

a. Thông tin cần phân tích:

Báo cáo giá trị AQI trung bình theo Bang, Loại chất lượng không khí (Category), DayLightSaving qua các năm.

b. Bối cảnh sự kiện -> dimension table:

- Ai : Không áp dụng
- Ở đâu: Các quận thuộc các bang tại Hoa Kỳ

- Cái gì: Phân loại chất lượng không khí (category_id, category_name, color_code, description)
- Khi nào: Thời gian ngày, tháng, năm và trạng thái DayLightSaving.

c. Đo lường (dữ kiện): Số lượng-> Fact Table

- Fact_AirQuality: Chứa các đo lường như AQI và thông tin về thời gian, vị trí, và loại chất lượng không khí.

d. Giá trị cần tính toán:

- DayLightSaving: Xác định dựa trên ngày tháng.

True: Từ 12/03/2023 đến 05/11/2023.

False: Ngoài khoảng trên.

- AQI trung bình: Tính trung bình theo Bang, Category và DayLightSaving qua các năm.

e. Phân cấp Chiều và Độ Mịn của Dữ Liệu:

- **Độ mịn dữ liệu (grain):** Mỗi dòng trong bảng sự kiện (Fact Table) tương ứng với một lần đo AQI tại một quận vào một ngày cụ thể.
- **Phân cấp chiều:**
 - Thời gian: Năm → Tháng → Ngày → DayLightSaving (True/False).
 - Địa điểm: Bang → Quận.
 - Loại AQI: Category.

1.9. Đếm số ngày theo Bang, Loại chất lượng không khí (Category) trong mỗi tháng. (bonus) (QUỐC)

Gợi ý phân tích:

- Loại chất lượng không khí trong bộ dữ liệu được tính cho từng **Quận**, không phải **Bang**.

a. Thông tin cần phân tích:

Số ngày theo từng loại chất lượng không khí

Phân tích theo bang và tháng

Tổng hợp từ dữ liệu cấp quận lên cấp bang

b. Bối cảnh sự kiện -> dimension table:

- Ai: Không có chiều người dùng
- Ở đâu: Dimension Geography (State > County)
- Cái gì: Dimension AQI_Category (phân loại chất lượng không khí)
- Khi nào: Dimension Time (Year > Month > Day)

c. Đo lường (dữ kiện): Số lượng-> Fact Table

- Số ngày xuất hiện của mỗi category
- AQI hằng ngày theo quận
- Count của số ngày theo category

d. Giá trị cần tính toán:

COUNT(Days) theo State, Category và Month

Tổng hợp từ County lên State

e. Phân cấp Chiều và Độ Mịn của Dữ Liệu:

- **Độ mịn dữ liệu (grain):** Mỗi dòng trong bảng sự kiện sẽ tương ứng với số ngày của một category trong một tháng của một bang
- **Phân cấp chiều:**
 - **Time: Year > Month > Day**
 - **Geography: State > County**
 - **AQI_Category: Category_Name > Range**

1.10. Báo cáo số ngày theo Loại chất lượng không khí (Category) và Tham số xác định (Defining Parameter). (bonus) (TÍN)

Gợi ý phân tích:

- Bạn có nhận xét gì về tình hình ô nhiễm không khí ở Hoa Kỳ nói chung?
- Thêm vào đó, hãy xác định các yếu tố chính mà quốc gia này cần xem xét để cải thiện chất lượng không khí.

a. Thông tin cần phân tích:

Báo cáo số ngày theo Loại chất lượng không khí (Category) và Tham số xác định (Defining Parameter).

b. Bối cảnh sự kiện -> dimension table:

- Ai : Không áp dụng
- Ở đâu: Không yêu cầu địa điểm cụ thể.
- Cái gì: Thông tin về loại chất lượng không khí (Good, Moderate, Unhealthy, etc.) và tham số xác định (Defining Parameter) như PM2.5, PM10, Ozone, CO, NO2.
- Khi nào: Thời gian đo chỉ số AQI.

c. Đo lường (dữ kiện): Số lượng-> Fact Table

- Fact_AirQuality: Ghi nhận số ngày thuộc từng loại chất lượng không khí và tham số xác định.

d. Giá trị cần tính toán:

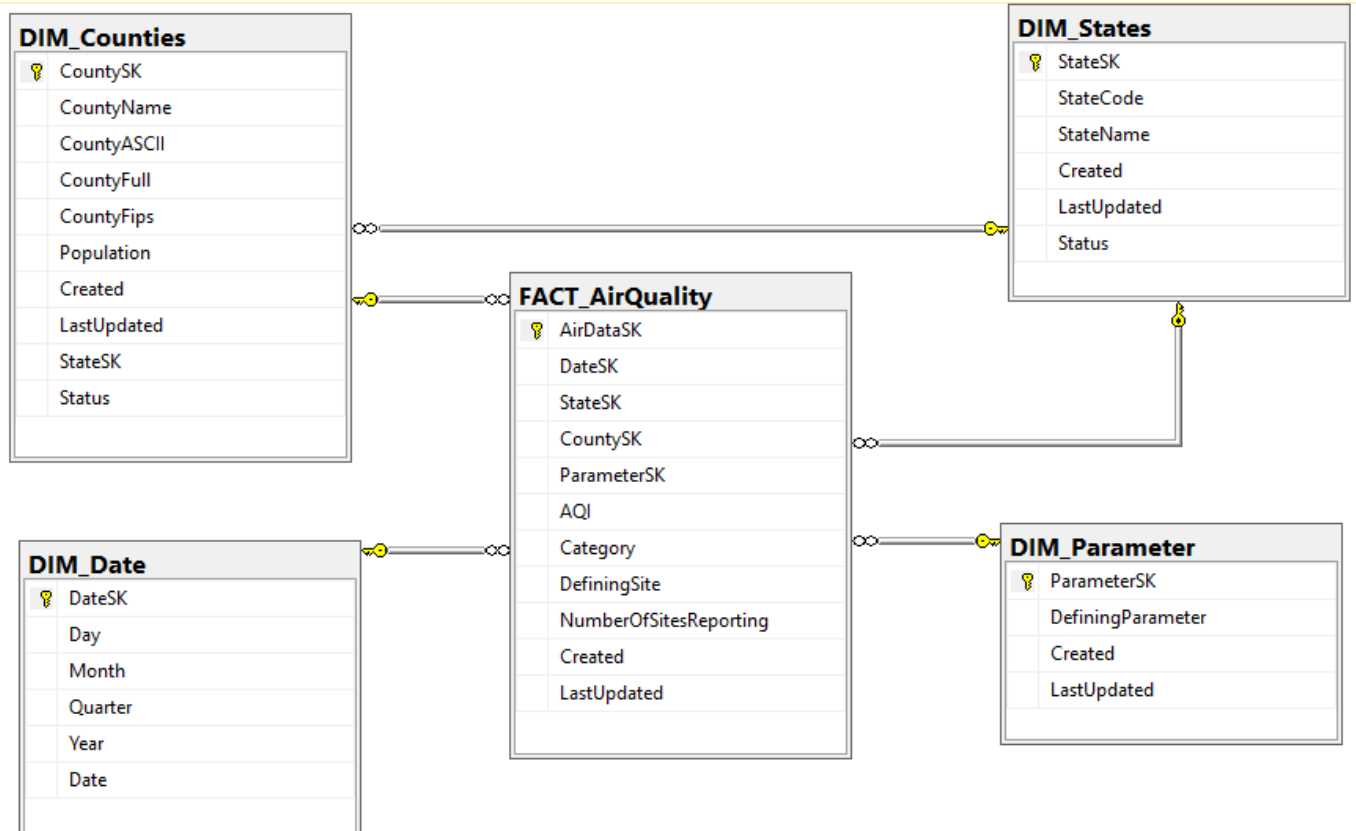
- Số ngày theo từng loại chất lượng không khí: Đếm số lượng ngày cho mỗi loại Category.
- Tham số xác định phổ biến nhất: Đếm số lần xuất hiện của mỗi tham số xác định và xác định tham số chiếm ưu thế.

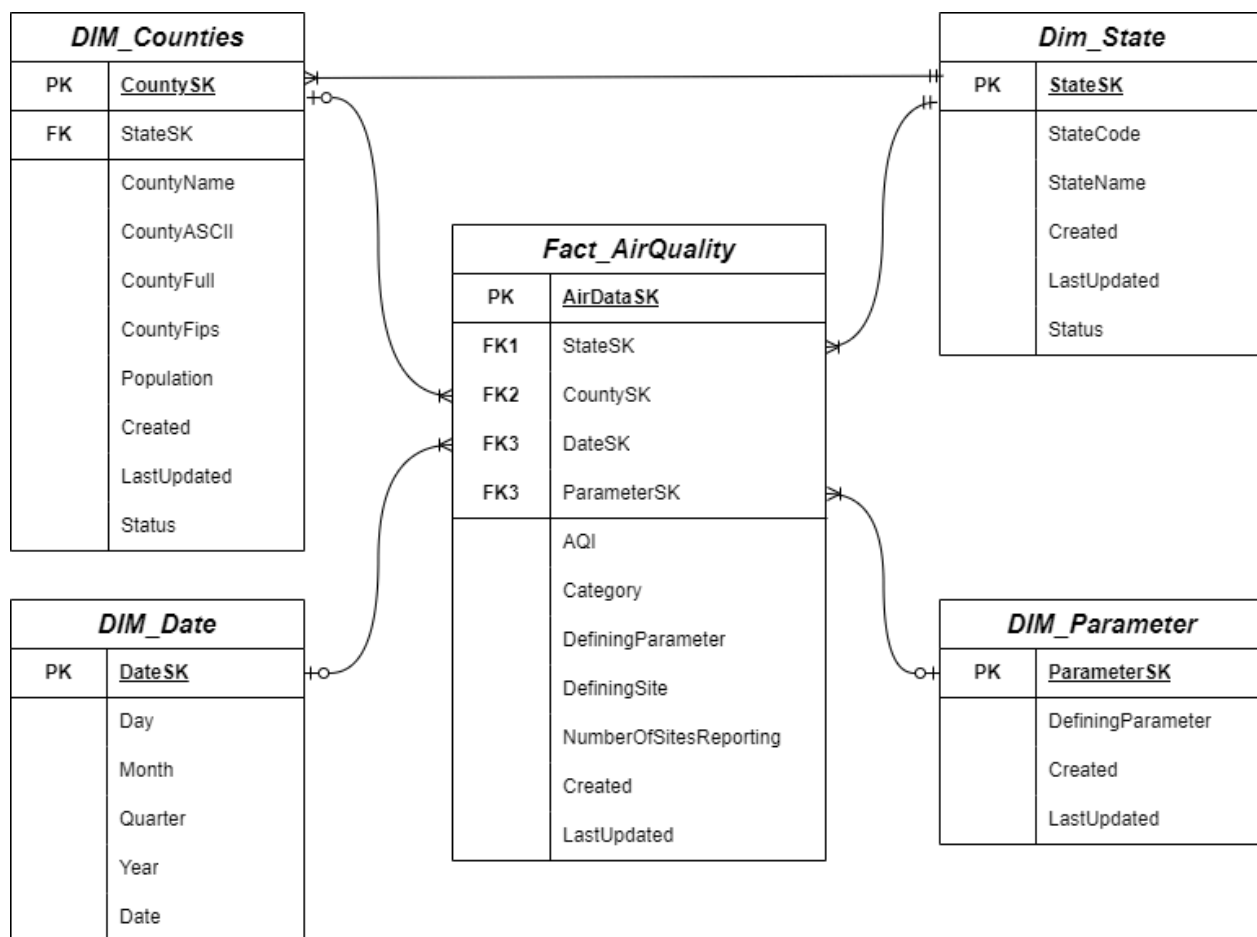
e. Phân cấp Chiều và Độ Mịn của Dữ Liệu:

- **Độ mịn dữ liệu (grain):** Mỗi dòng trong Fact Table sẽ tương ứng với một ngày đo AQI và tham số xác định cụ thể.
- **Phân cấp chiều:**

- Thời gian: Năm → Tháng → Ngày.
- Loại chất lượng không khí: Good → Moderate → Unhealthy → Hazardous.

2) Cấu trúc bảng DDS:





a. Bảng DIM_States:

STT	Tên thuộc tính	Mô tả	Transformation rules	Nguồn	SCD
1	StateSK	Khóa thay thế duy nhất cho mỗi stage	Khóa tự tăng được gán trong quá trình ETL		
2	StateCode	Mã của bang (Natural Key)		NDS_States	2
3	StateName	Mã của bang		NDS_States	2
4	Created	Tên của bang		NDS_States	
5	LastUpdated	Ngày tạo bản ghi		NDS_States	
6	Status	Trạng thái của bản ghi (hoạt động)	Gán giá trị 1 (Bản ghi hiện tại) hoặc 0 (Bản ghi cũ)	ETL Process	

b. Bảng DIM_Counties:

STT	Tên thuộc tính	Mô tả	Transformation rules	Nguồn	SCD
1	CountySK	Khóa thay thế duy nhất cho mỗi County	Khóa tự tăng được gán trong quá trình ETL	NDS_Counties	
2	CountyName	Tên của quận		NDS_Counties	2
3	CountyASCII	Tên ASCII của quận		NDS_Counties	2
4	CountyFull	Tên đầy đủ của quận		NDS_Counties	2
5	CountyFips	Mã FIPS của quận (Natural Key)		NDS_Counties	
6	Population	Dân số của quận		NDS_Counties	1
7	Created	Ngày tạo bản ghi		NDS_Counties	
8	LastUpdated	Ngày cập nhật bản ghi		NDS_Counties	
9	StateSK	Khóa thay thế của Stage mà Counties thuộc	Khóa ngoại từ NDS_Counties	NDS_Counties	
10	Status	Trạng thái của bản ghi (hoạt động)	Gán giá trị 1 (Bản ghi hiện tại) hoặc 0 (Bản ghi cũ)	ETL Process	

c. Bảng DIM_Date:

STT	Tên thuộc tính	Mô tả	Transformation rules	Nguồn	SCD
1	DateSK	Khóa thay thế duy nhất cho mỗi Date	Khóa tự tăng được gán trong quá trình ETL	ETL Process	
2	Day	Ngày		ETL Process	

3	Month	Tháng		ETL Process	
4	Quarter	Quý		ETL Process	
5	Year	Năm		ETL Process	
6	Date	Ngày đo		NDS_AirData	

d. Bảng DIM_Parameter:

STT	Tên thuộc tính	Mô tả	Transformation rules	Nguồn	SCD
1	ParameterSK	Khóa thay thế duy nhất cho mỗi Date	Khóa tự tăng được gán trong quá trình ETL	ETL Process	
2	DefiningParameter	Tên hoặc giá trị của tham số chất ô nhiễm		NDS_AirData	
3	Created	Ngày tạo bản ghi		NDS_AirData	
4	LastUpdated	Ngày cập nhật bản ghi		NDS_AirData	


e. Bảng FACT_AirQuality:

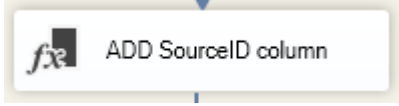
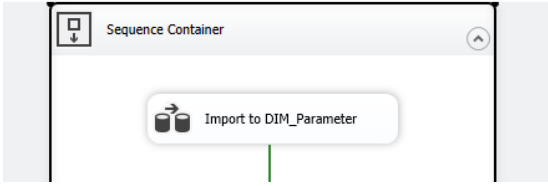
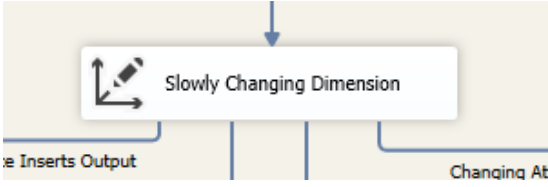
STT	Tên thuộc tính	Mô tả	Transformation rules	Nguồn	SCD
1	AirDataSK	Khóa thay thế duy nhất cho mỗi Date	Khóa tự tăng được gán trong quá trình ETL	ETL Process	
2	DateSK	Khóa thay thế cho ngày, liên kết với bảng DIM_Dates	Khóa ngoại từ NDS_AirData	ETL Process	
3	StateSK	Khóa thay thế của County mà AirData thuộc	Khóa ngoại từ NDS_AirData	NDS_AirData	
4	CountySK	Khóa thay thế của Stage mà	Khóa ngoại từ NDS_AirData	NDS_AirData	

		County thuộc			
5	ParameterSK	Khóa thay thế đại diện cho thông số đo lường, liên kết với bảng DIM_Parameter	Khóa ngoại từ NDS_AirData	ETL Process	
6	AQI	Loại chỉ số AQI		NDS_AirData	
7	Category	Thông số xác định		NDS_AirData	
8	DefiningParameter	Mã của địa điểm xác định		NDS_AirData	
9	DefiningSite	Số lượng địa điểm báo cáo		NDS_AirData	
10	NumberOfSitesReporting	Ngày tạo bản ghi		NDS_AirData	
11	Created	Ngày cập nhật bản ghi		NDS_AirData	
12	LastUpdated	Loại chỉ số AQI		NDS_AirData	


3) Data Flow:

Danh sách các transformation đã sử dụng

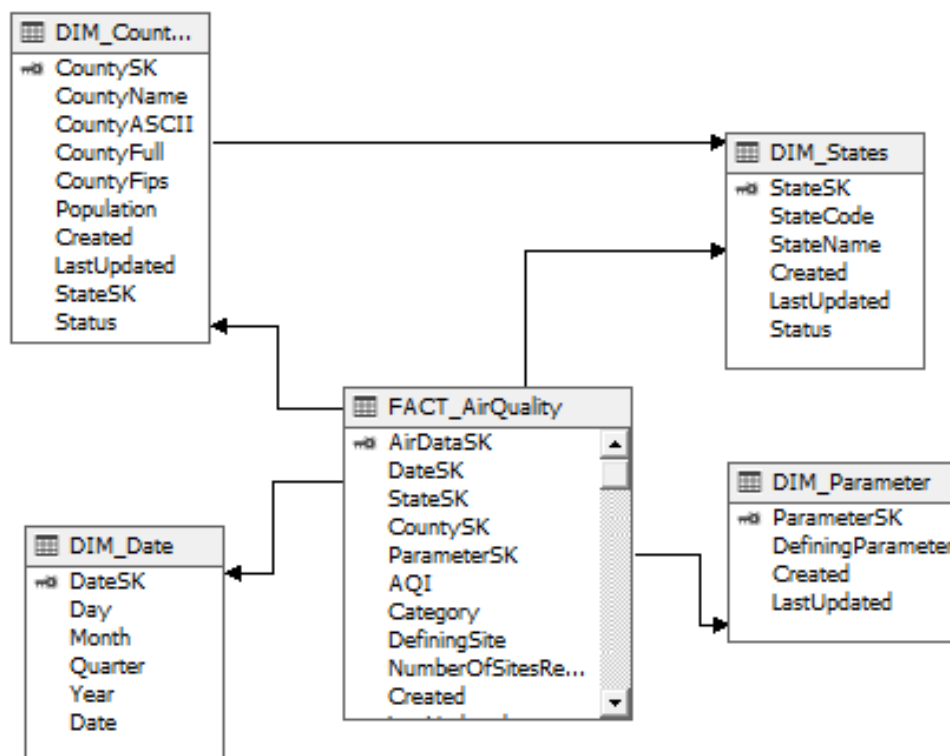
Transformation	Mục đích sử dụng	Ghi chú
Lookup	Tra cứu dữ liệu từ bảng khác và thêm giá trị cột mới. (Sử dụng để liên kết với bảng Stage lấy khóa NK hoặc tạo khóa ngoại trong NDS)	

Derived Column	Tính toán giá trị mới hoặc sửa đổi giá trị cột.	
Sequence Container	Nhóm các tác vụ thành một đơn vị logic, quản lý luồng công việc, xử lý lỗi và tối ưu hóa quy trình ETL.	
Slowly Changing Dimension	Quản lý thay đổi dữ liệu theo thời gian trong các bảng dimension, giúp theo dõi các thay đổi như thêm, sửa, hoặc xóa thông tin mà không làm mất dữ liệu lịch sử.	

II. PHÂN TÍCH QUY TRÌNH ETL:

Link NDS-> DDS:  [BI2425] HCMUS ETL NDS to DDS

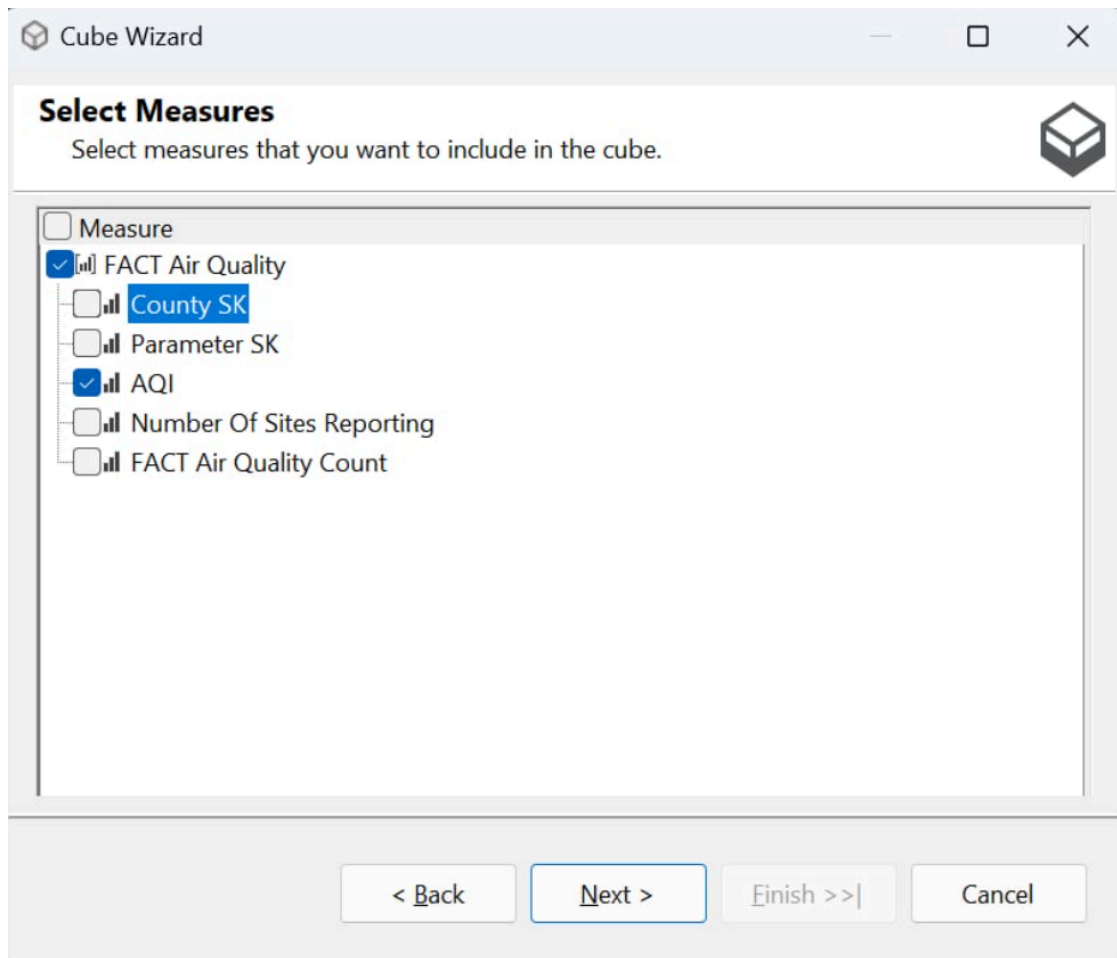
III. GIAI ĐOẠN 3: OLAP:

*Data Source View*

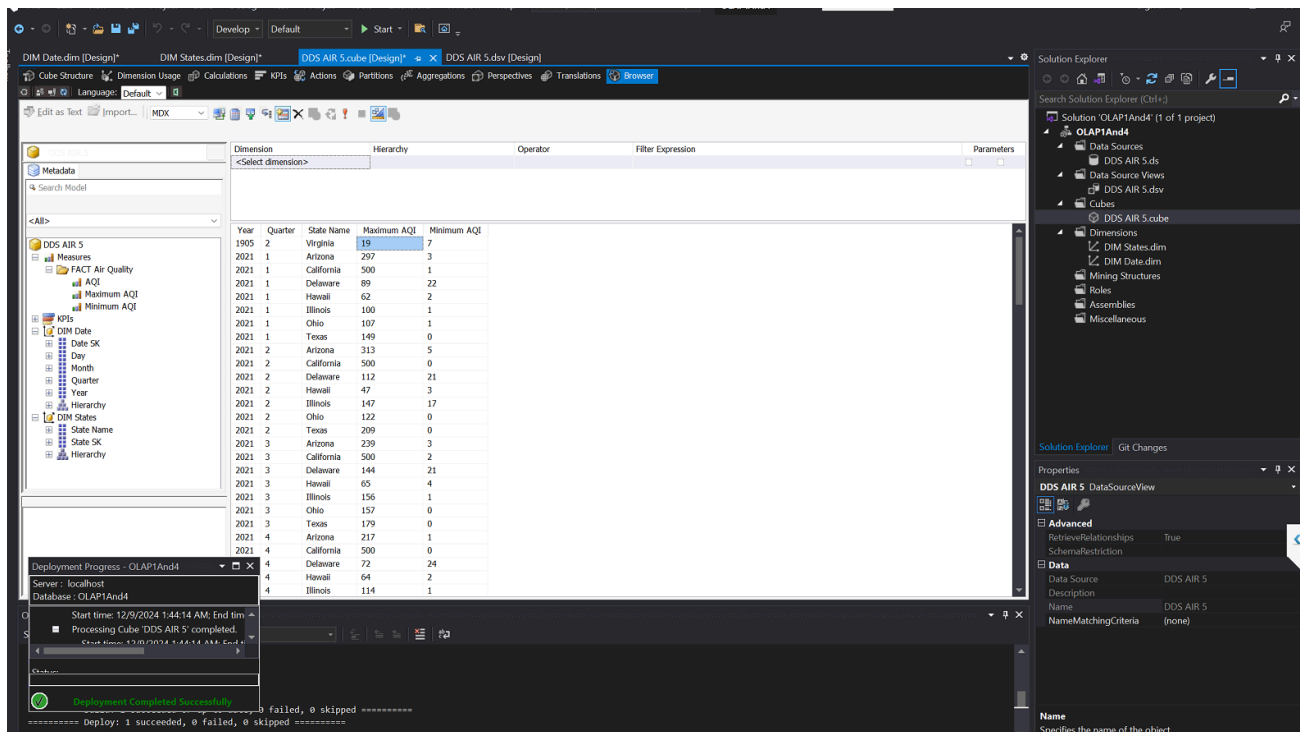
1.1 Báo cáo giá trị AQI nhỏ nhất và lớn nhất cho từng bang trong mỗi quý của các năm.

Để thực hiện phân tích AQI theo từng bang và theo quý năm, chúng ta cần truy cập các bảng dữ liệu sau:

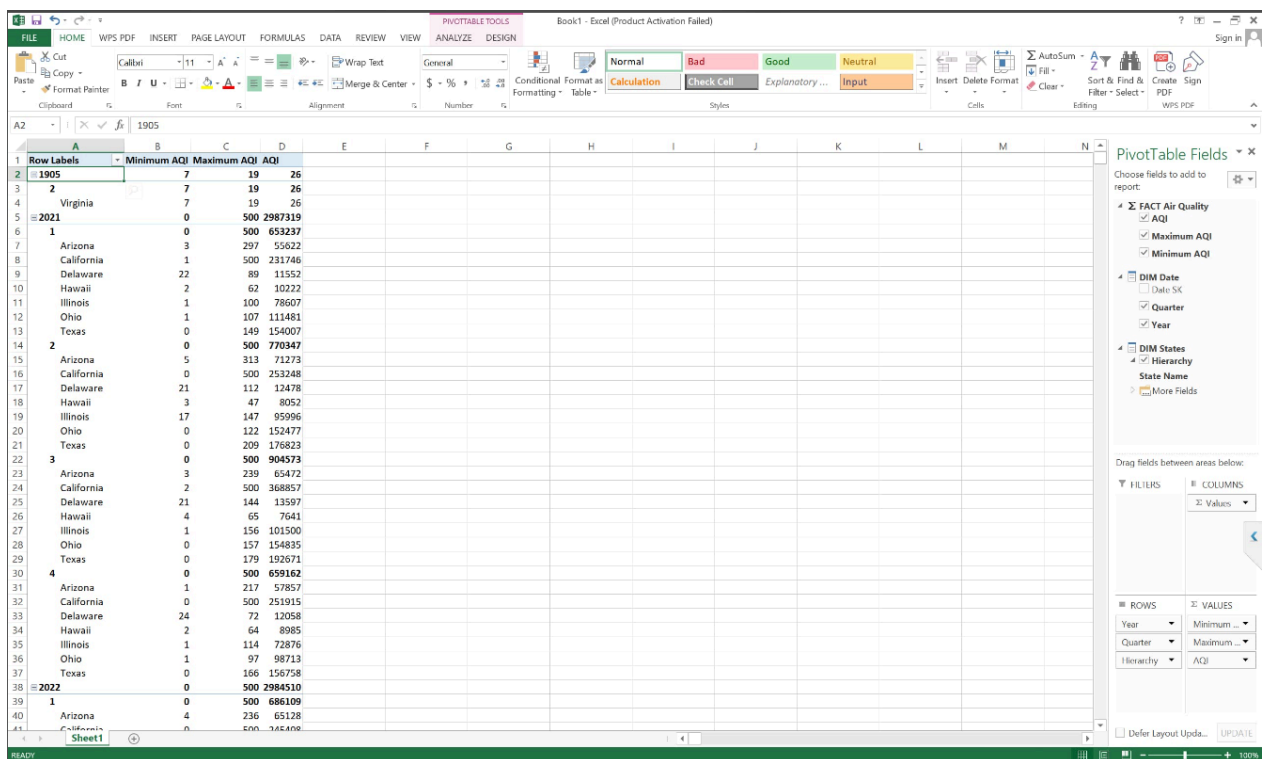
- **Bảng fact_AirQuality:** Cung cấp các giá trị AQI min và max theo từng ngày.
- **Bảng Dim States:** Cung cấp tên các bang để phân tích theo từng bang.
- **Bảng Dim Date:** Cung cấp thông tin về quý (quarter) và năm (year) để phân tích theo từng quý trong các năm.



- **Truy xuất dữ liệu AQI:** Sử dụng bảng **fact_AirQuality** để lấy giá trị AQI min và max.
- **Xây dựng Cube:** Cube sẽ có các chiều từ bảng **Dim States** (tên bang) và **Dim Date** (quý và năm) để phân loại dữ liệu.
- **Deploy Cube:** Triển khai Cube và thiết lập các hierarchy cho **Dim Date** và **Dim States**.
- **Tạo Measurements:** Sử dụng hàm **min** và **max** để tính toán mà không cần MDX.
- **Kiểm tra kết quả:** Kéo thả các trường cần thiết vào bảng trong **Browser** và nhấn **Execute** để kiểm tra kết quả.

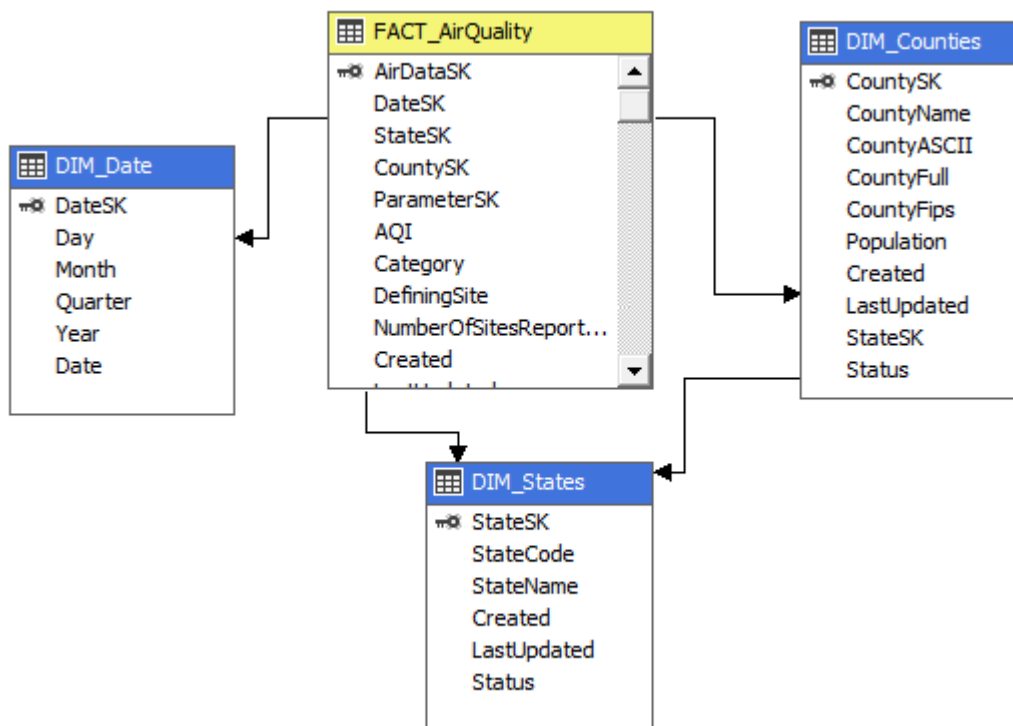


Vào **Data**, chọn **Open Existing Connection** và kết nối với server Cube. Kéo thả các trường dữ liệu cần thiết (AQI min, max, năm, quý, bang)

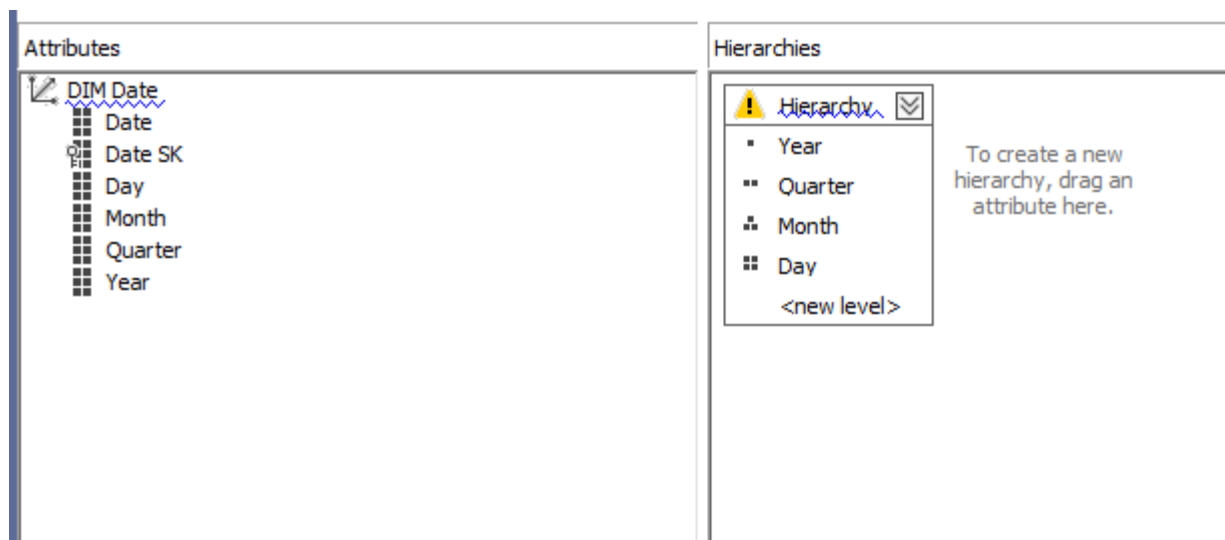


1.3. Báo cáo số ngày (number of days) và giá trị AQI trung bình (mean AQI) khi chất lượng không khí được đánh giá là "rất có hại" (very unhealthy) hoặc tồi tệ hơn cho từng bang (State) và từng quận (County).

a. Cubes:



b. Phân cấp các chiều dữ liệu:



Phân cấp chiều DIM_Date

c. Measures:

- Xác định số ngày (**Number Of Days**): Sử dụng hàm COUNT để đếm số ngày từ [Dim Date].[Date].[Date] mà chất lượng không khí thuộc các mức "Very Unhealthy" hoặc "Hazardous".
- Tính chỉ số AQI trung bình (**Mean AQI**): Sử dụng hàm AVG để tính trung bình chỉ số AQI cho những ngày được lọc bởi điều kiện chất lượng không khí.

d. Kết quả phân tích:

Messages			Results	
			Number Of Days	MeanAQI
All	Bexar	Hazardous	1	15
All	Butte	Very Unhealthy	3	244.333333333333
All	Champaign	Very Unhealthy	1	40
All	Coconino	Very Unhealthy	1	38
All	Colusa	Very Unhealthy	1	225
All	Cook	Very Unhealthy	2	41
All	Cuyahoga	Very Unhealthy	1	55
All	Dunn	Hazardous	1	48
All	DuPage	Very Unhealthy	2	29.5
All	El Dorado	Hazardous	1	375
All	El Dorado	Very Unhealthy	1	274
All	Franklin	Very Unhealthy	1	61
All	Fresno	Very Unhealthy	2	250.5
All	Harris	Very Unhealthy	1	23
All	Humboldt	Very Unhealthy	4	107.75
All	Imperial	Hazardous	12	134.583333333333
All	Imperial	Very Unhealthy	6	91.6666666666667
All	Inyo	Hazardous	14	144.285714285714
All	Inyo	Very Unhealthy	8	134.25

1.4 Đếm số ngày trong từng loại chất lượng không khí (Good, Moderate, Unhealthy, etc.) cho mỗi quận (County) thuộc bốn bang: Hawaii, Alaska, Illinois, và Delaware.

a. Xác định Cube:

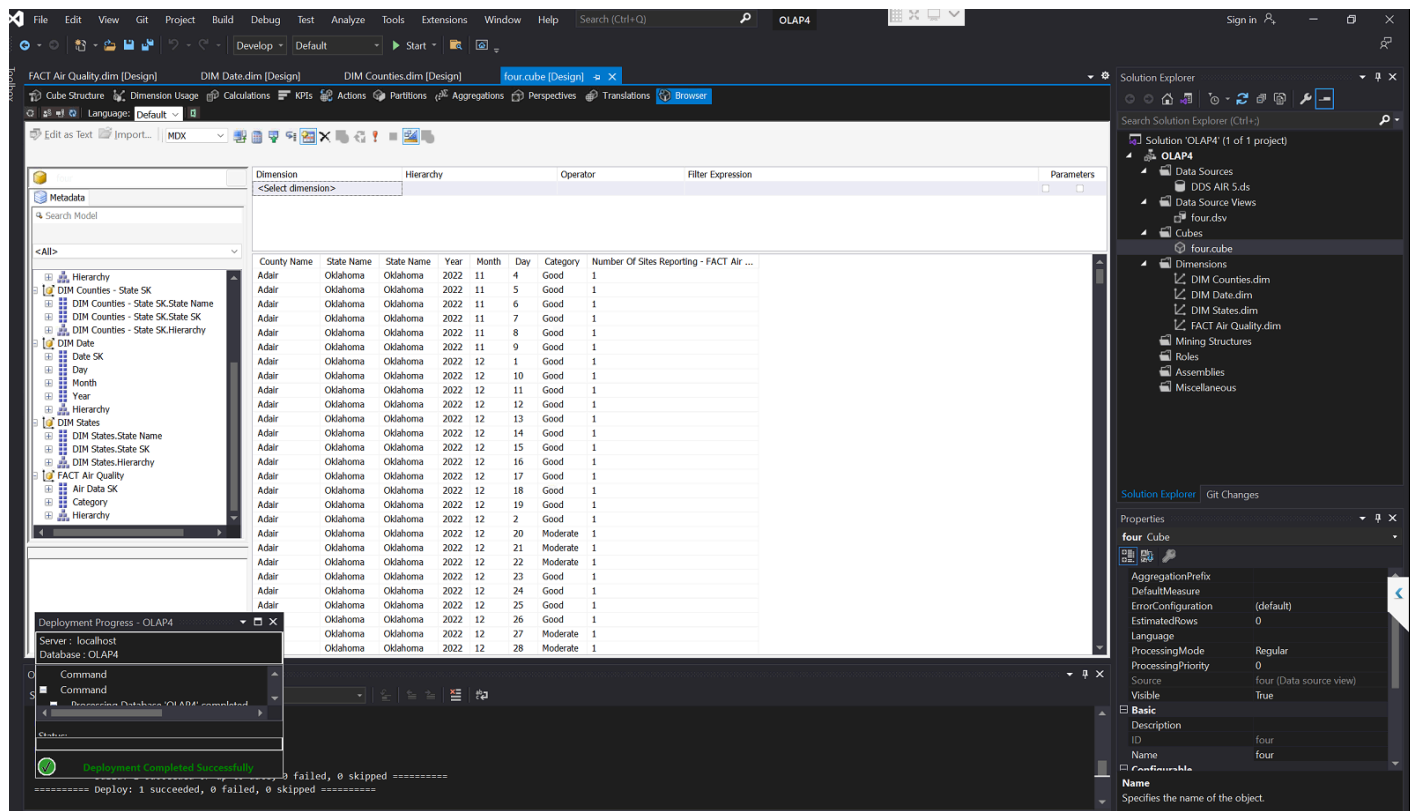
- Cube chính: **fact_AirQuality** để chứa dữ liệu AQI và **NumberOfSiteReporting**.

b. Xác định các bảng dim:

- **dim_date**: Để đếm số ngày theo yêu cầu.
- **dim_country**: Để xác định từng quận.
- **dim_state**: Để lọc theo 4 bang.

c. Các bước thực hiện:

- Sử dụng hàm COUNT hoặc SUM tùy theo yêu cầu.
- Tích hợp các thông số cần thiết để triển khai, đảm bảo dữ liệu đáp ứng đúng yêu cầu bài toán.



Tích hợp đầy đủ các bảng như **fact_AirQuality**, **dim_date**, **dim_country**, và **dim_state** vào Excel, giống như trong câu 1. Bảng dữ liệu cuối cùng chỉ hiển thị thông tin liên quan

đến 4 bảng được chọn, đáp ứng đúng yêu cầu bài toán.

FILEHOMEWPS PDFINSERTPAGE LAYOUTFORMULASDATA REVIEWVIEWANALYZEDESIGN

CutCopyFormat PainterClipboard

Calibri11

B

I

U

Font

Wrap Text

Merge & Center

Alignment

General

\$ % & #

Number

Conditional Formatting

Format as Table

NormalBadGoodNeutral

CalculationCheck CellExplanatory...

Input

AutoSum

Fill

Clear

InsertDelete Format

Columns

Sort & Filter

Find & Select

Create PDF

Sign

WPS PDF

Book3 - Excel (Product Activation Failed)

Sign in

1234567891011121314151617181920212223242526272829303132333435363738394041

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

Delaware Total

Hawaii

Very Unhealthy

Good

Moderate

Unhealthy

Unhealthy for Sensitive Groups

403

1228

960

4589

846

623

647

664

966

807

561

511

73

3035

577

582

1242

682

583

1035

981

975

16

72

860

67

10682

897

85

434

401

83

561

143

73

1021

387

480

710

1124

765

1021

387

480

710

1124

765

2

111

6

131

6

1

1

1

10

2

8

8

8

4

4

2

3

3

6

2

167

12

26

6

880

28

10

12

24

8

34

24

12

138

21

13

48

14

19

42

36

34

1460

33

4

33

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

53

487

2118

1033

16295

1781

718

1094

9890

4205

1090

1058

1412

1692

730

1091

1086

4433

1744

1093

1095

6105

2313

1085

1086

1792

1143

2151

1780

46821

72054

487

2118

1033

16295

1781

718

1094

9890

4205

1090

1058

1412

1692

730

1091

1086

4433

1744

1093

1095

6105

2313

1085

1086

1792

1143

2151

1780

46821

72054

PivotTable Fields

Choose fields to add to report

DIM Countries

State SK

DIM Counties

State

More Fields

DIM Date

Hierarchy

Year

Month

Day

More Fields

DIM States

DIM States.Hiera...

More Fields

FACT Air Quality

Hierarchy

More Fields

Drag fields between areas below:

FILTERS

Hierarchy

COLUMNS

DIM State...

Hierarchy

ROWS

Hierarchy

VALUES

Number O...

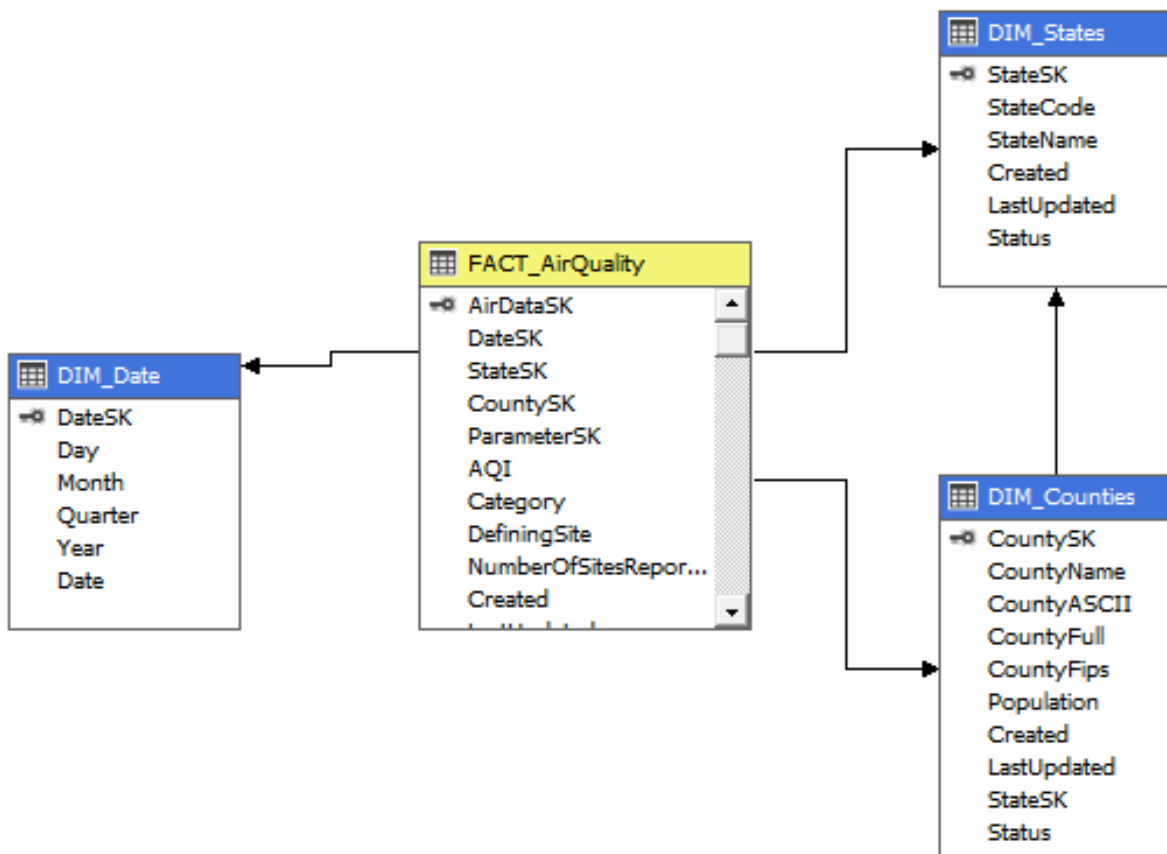
Defer Layout Update

UPDATE

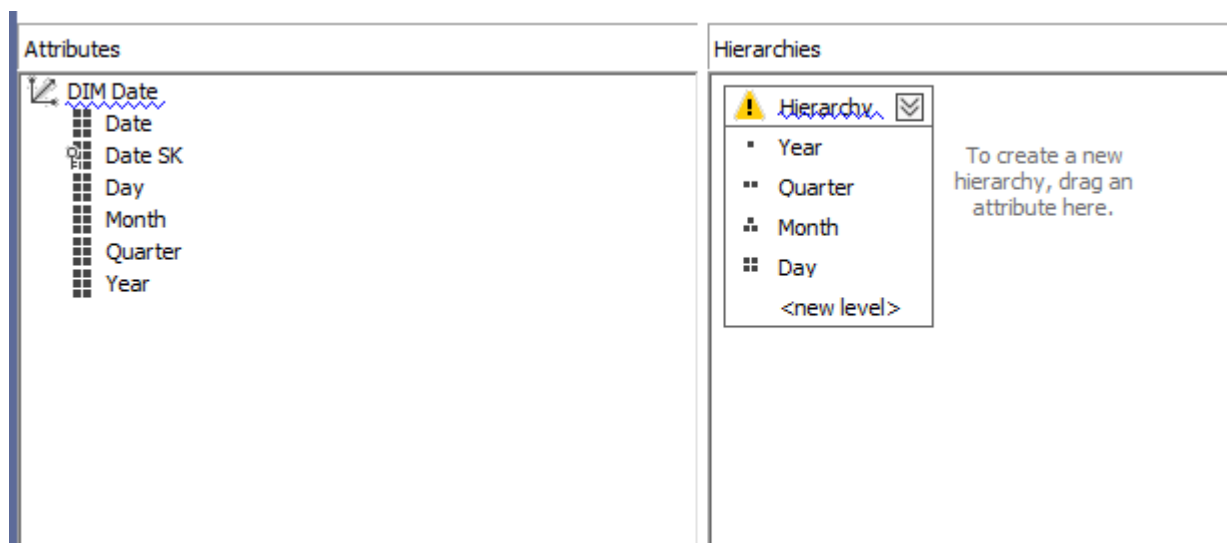
Messages			Results
			Number of Days
Delaware	Kent	All	1072
Delaware	Kent	Good	753
Delaware	Kent	Moderate	309
Delaware	Kent	Unhealthy for Sensitive Groups	7
Delaware	Kent	Unhealthy	2
Delaware	Kent	Very Unhealthy	1
Delaware	New Castle	All	1095
Delaware	New Castle	Good	565
Delaware	New Castle	Moderate	516
Delaware	New Castle	Unhealthy for Sensitive Groups	10
Delaware	New Castle	Unhealthy	2
Delaware	New Castle	Very Unhealthy	2
Delaware	Sussex	All	606
Delaware	Sussex	Good	449
Delaware	Sussex	Moderate	157
Hawaii	Hawaii	All	1088
Hawaii	Hawaii	Good	993
Hawaii	Hawaii	Moderate	95
Hawaii	Honolulu	All	1095
Hawaii	Honolulu	Good	1057
Hawaii	Honolulu	Moderate	37
Hawaii	Honolulu	Unhealthy for Sensitive Groups	1

1.6. Thiết kế một báo cáo để thể hiện xu hướng biến động của AQI trong năm cho bốn bang: Hawaii, Alabama, Utah, và Georgia.

a. Cubes:



b. Phân cấp các chiều dữ liệu:



Phân cấp chiều DIM_Date

c. Measures:

- Tính giá trị trung bình AQI (**MeanAQI**) cho từng tháng: Sử dụng các thành viên trong phần [Dim Date].[Date] (ngày) để giới hạn dữ liệu trong một tháng nhất định.
- $AVG([Measures].[AQI])$: Tính giá trị trung bình của chỉ số AQI cho mỗi tháng từ các giá trị AQI của các ngày trong tháng đó.

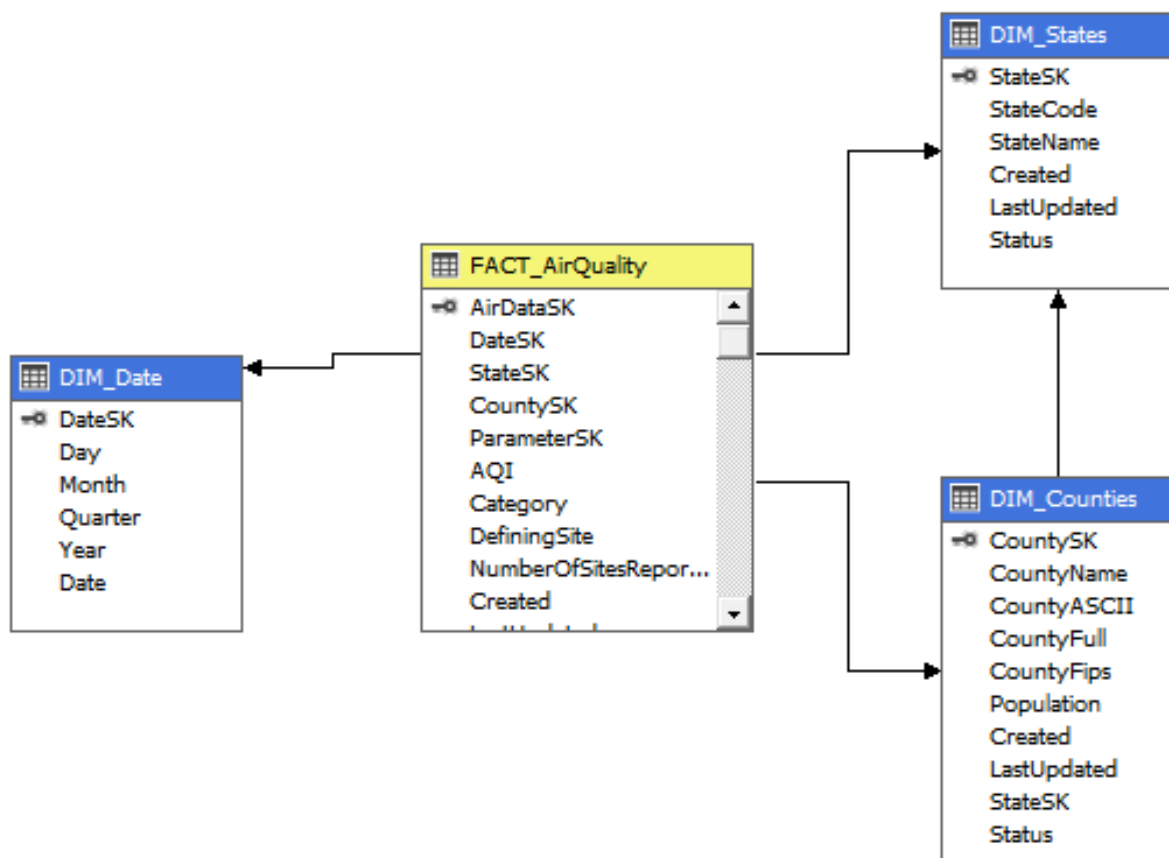
d. Kết quả phân tích:

121 %

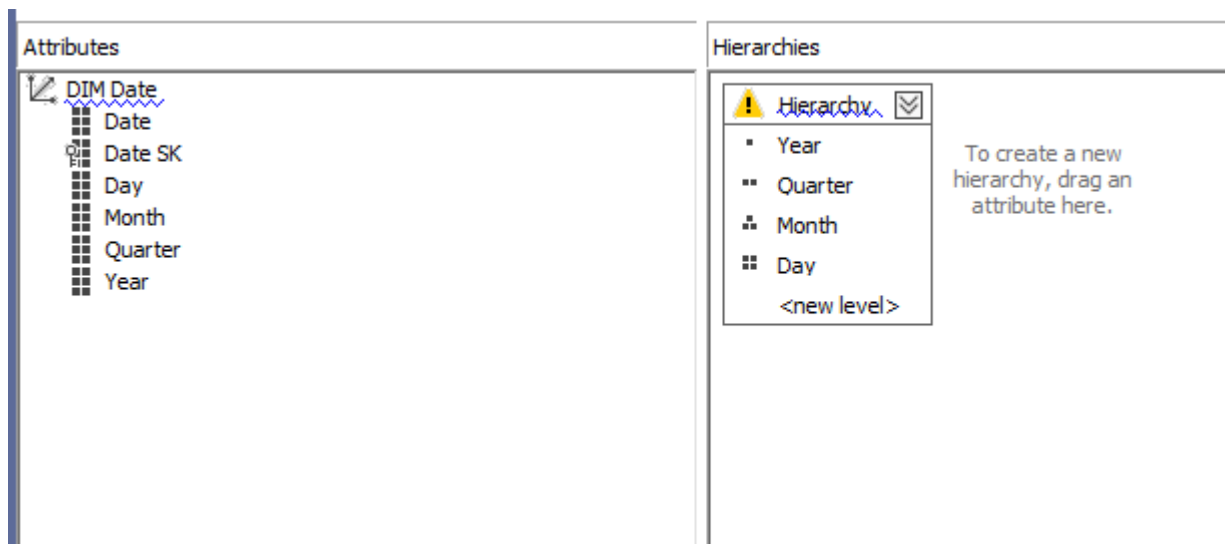
	Hawaii	Alabama	Utah	Georgia
1	97.5913978494624	16.5161290322581	149.258064516129	30.3214285714286
10	82.247311827957	20.3114754098361	133.274193548387	37.1290322580645
11	81.6111111111111	19.25	137.283333333333	52.7
12	81.8817204301075	34.375	129.693548387097	54.9354838709677
2	95.0476190476191	18.9642857142857	140.178571428571	33.3461538461538
3	94.5591397849462	23.5806451612903	166.91935483871	29.5806451612903
4	77.1555555555556	39.6	168.333333333333	26
5	75.2903225806452	28.6774193548387	159.870967741935	27.7096774193548
6	64.6666666666667	27.8333333333333	123.9	25.8333333333333
7	67.0430107526882	40.2903225806452	122.435483870968	23.6129032258065
8	65.8817204301075	22.7678571428571	132.629032258065	38.7096774193548
9	65.2666666666667	17.8833333333333	126.05	38.9

**1.7. Báo cáo trung bình, độ lệch chuẩn, giá trị nhỏ nhất (min) và giá trị lớn nhất (max) của chỉ số AQI, nhóm theo Bang và Quận trong mỗi quý của năm.
(bonus)**

a. Cubes:



b. Phân cấp các chiều dữ liệu:



Phân cấp chiều DIM_Date

c. Measures:

- AQI Average: Tính giá trị trung bình của AQI cho từng kết hợp giữa bang, quận, năm và quý.
- AQI StdDev: Tính độ lệch chuẩn của AQI cho từng kết hợp trên.
- AQI Min: Tính giá trị AQI thấp nhất.
- AQI Max: Tính giá trị AQI cao nhất.

d. Kết quả phân tích:

Messages				Results			
				AQI Average	AQI StdDev	AQI Min	AQI Max
Alabama	Bartholomew	1905	2	4112.4676159853	1814.80012116051	7	14793
Alabama	Bartholomew	1905	3	4112.4676159853	1814.80012116051	7	14793
Alabama	Bartholomew	1905	4	4112.4676159853	1814.80012116051	7	14793
Alabama	Bartholomew	2021	1	4112.4676159853	1814.80012116051	7	14793
Alabama	Bartholomew	2021	2	4112.4676159853	1814.80012116051	7	14793
Alabama	Bartholomew	2021	3	4112.4676159853	1814.80012116051	7	14793
Alabama	Bartholomew	2021	4	4112.4676159853	1814.80012116051	7	14793
Alabama	Bartholomew	2022	1	4112.4676159853	1814.80012116051	7	14793
Alabama	Bartholomew	2022	2	4112.4676159853	1814.80012116051	7	14793
Alabama	Bartholomew	2022	3	4112.4676159853	1814.80012116051	7	14793
Alabama	Bartholomew	2022	4	4112.4676159853	1814.80012116051	7	14793
Alabama	Bartholomew	2023	1	4112.4676159853	1814.80012116051	7	14793
Alabama	Bartholomew	2023	2	4112.4676159853	1814.80012116051	7	14793
Alabama	Bartholomew	2023	3	4112.4676159853	1814.80012116051	7	14793
Alabama	Bartholomew	2023	4	4112.4676159853	1814.80012116051	7	14793
Alabama	Bell	1905	2	4112.4676159853	1814.80012116051	7	14793

PHÂN CÔNG CÔNG VIỆC

STT	Công việc	Người thực hiện	Ngày hoàn thành	Hoàn thành (%)
GIAI ĐOẠN 1: DDS				
1	Tạo template Báo cáo, phân công	Lô Thủy Tiên	00:00 AM 27/11/2024	100%
2	<i>Phân tích yêu cầu DDS:</i> <ul style="list-style-type: none"> - 1.1. Báo cáo giá trị AQI nhỏ nhất và lớn nhất cho từng bang trong mỗi quý của các năm. - 1.5. Tính giá trị trung bình AQI theo từng quý cho bốn bang: Hawaii, Alaska, Illinois, và Delaware. - 1.11. Đếm số ngày theo Bang, Loại chất lượng không khí (Category) trong mỗi tháng. (bonus) 	Võ Lâm Hải Quốc	9:00 AM 29/11/2024	100%
3	<i>Phân tích yêu cầu DDS:</i> <ul style="list-style-type: none"> - 1.2. Báo cáo giá trị trung bình (mean) và độ lệch chuẩn (standard deviation) của chỉ số AQI cho từng bang trong mỗi quý của các năm. - 1.9. Báo cáo trung bình, độ lệch 	Nguyễn Vũ Tường An	9:00 AM 29/11/2024	100%

	chuẩn, giá trị nhỏ nhất (min) và giá trị lớn nhất (max) của chỉ số AQI, nhóm theo Bang và Quận trong mỗi quý của năm. (bonus)			
4	<i>Phân tích yêu cầu DDS:</i> <ul style="list-style-type: none"> - 1.3. Báo cáo số ngày (number of days) và giá trị AQI trung bình (mean AQI) khi chất lượng không khí được đánh giá là "rất có hại" (very unhealthy) hoặc tồi tệ hơn cho từng bang (State) và từng quận (County). - 1.6. Thiết kế một báo cáo để thể hiện xu hướng biến động của AQI trong năm cho bốn bang: Hawaii, Alaska, Illinois, và California. - 1.12. Báo cáo số ngày theo Loại chất lượng không khí (Category) và Tham số xác định (Defining Parameter). (bonus) 	Võ Trung Tín	9:00 AM 29/11/2024	100%
5	<i>Phân tích yêu cầu DDS:</i> <ul style="list-style-type: none"> - 1.4. Đếm số ngày trong từng loại chất lượng không khí (Good, Moderate, Unhealthy, etc.) cho mỗi quận (County) thuộc bốn bang: Hawaii, Alaska, Illinois, 	Lô Thủy Tiên	9:00 AM 29/11/2024	100%

	và Delaware. - 1.10. Tạo một thuộc tính mới, DayLightSaving, trong bảng phù hợp. DayLightSaving có hai giá trị: (bonus)			
6	Cấu trúc bảng DDS: - Thiết kế script SQL DDS - Thiết kế diagram DDS	Lô Thủy Tiên	9:00 AM 01/12/2024	100%
GIAI ĐOẠN 2: NDS-> DDS				
3	Xây dựng SSIS package để load data NDS-> DDS	Lô Thủy Tiên	00:00 AM 05/12/2024	100%
7	Xây dựng SSIS package	Võ Lâm Hải Quốc Nguyễn Vũ Tường An Võ Trung Tín (Tham gia google meet)	00:00 AM 07/12/2024	50%
GIAI ĐOẠN 3: OLAP				
8	- OLAP, tạo truy vấn, visual 1.1, 1.4	Võ Lâm Hải Quốc	00:00 AM 09/12/2024	100%
9	- OLAP, tạo truy vấn, visual 1.2, 1.5	Nguyễn Vũ Tường An	00:00 AM 09/12/2024	0%

10	- OLAP, tạo truy vấn, visual 1.3, 1.6	Võ Trung Tín	00:00 AM 09/12/2024	100%
11	- OLAP, tạo truy vấn, visual 1.7	Lô Thủy Tiên	00:00 AM 09/12/2024	100%
12	Hoàn thành báo cáo tổng hợp GD2: ETL NDS-> DDS, OLAP	Lô Thủy Tiên	00:00 AM 10/12/2024	100%
13	<i>Video giải thích từng component trong ETL process</i>	Lô Thủy Tiên	00:00 AM 10/12/2024	100%
14	<i>Video giải thích OLAP, tạo truy vấn, visual : 1.1, 1.4</i>	Võ Lâm Hải Quốc	00:00 AM 10/12/2024	100%

Báo cáo từ Github:

Link: <https://github.com/Locung201/BI2425>

Demo:

Link NDS-> DDS:  [BI2425] HCMUS ETL NDS to DDS

Link OLAP:

https://drive.google.com/drive/folders/1DX5nIt82hp0_mkCbrZhPVJydDMM40xQx?fbclid=IwZXh0bgNhZW0CMTEAAR3rUeQ3ce6AzhuPvDod-xYZy-ex9r67f3_q2t77XmwVdnGcKyPVy8oU4i8_aem_k8_gvZt44OiSUCsP9WSwFw

ĐÁNH GIÁ THÀNH VIÊN

Nhóm : 03

MSSV	Họ tên	Hoàn thành	Nhận xét
19127536	Võ Lâm Hải Quốc	85%	
21127211	Nguyễn Vũ Tường An	30%	Trễ deadline
21127450	Võ Trung Tín	80%	
21127699	Lô Thủy Tiên	100%	

TÀI LIỆU THAM KHẢO**Công cụ và phần mềm hỗ trợ:**

STT	Chức năng	Công cụ
[1]	Thiết kế SSIS	Microsoft Visual Studio 2022
[2]	Báo cáo	Google Docs
[3]	Quản lý, trao đổi	Facebook, Messenger
[4]	Họp định kỳ	Google Meet
[5]	Quay Video demo	OBS Open Broadcaster Software
[6]	Thuyết trình	Canva
[7]	Sản phẩm Demo	Youtube
[8]	AI support	ChatGPT, Copilot, Perplexity, Gemini
[9]	Quản lý dự án	SSMS 20
[10]	Thiết kế CSDL	SQL server

Tài liệu tham khảo:

- [0] Tài liệu môn học Hệ thống thông tin phục vụ trí tuệ kinh doanh - 21HTTT2
- [1] Microsoft. (2024). *SQL Server Integration Services*.
<https://learn.microsoft.com/en-us/sql/integration-services/sql-server-integration-services?view=sql-server-ver16>
- [2] Đông Chí . (24 thg 12, 2019). *Phần 2: Đổ dữ liệu từ NDS sang DDS*
YouTube.<https://www.youtube.com/watch?v=128Oldh5uMo>.
- [3] Chuc Nguyen Van. (3 thg 7, 2021). *ETL Project From Excel Data Source to Star Schema with SSIS*. YouTube.<https://www.youtube.com/watch?v=Yp8fXLnVCp8&t=597s>
- [4] Huy Bui. (2022, 30 tháng 9). *SQL Server Integration Services – SSIS*. Cole.edu.vn.
<https://cole.edu.vn/sql-server-integration-services-ssis/>
- [5] Microsoft. (n.d.). *Microsoft Learn*. Link: <https://learn.microsoft.com/en-us/> .