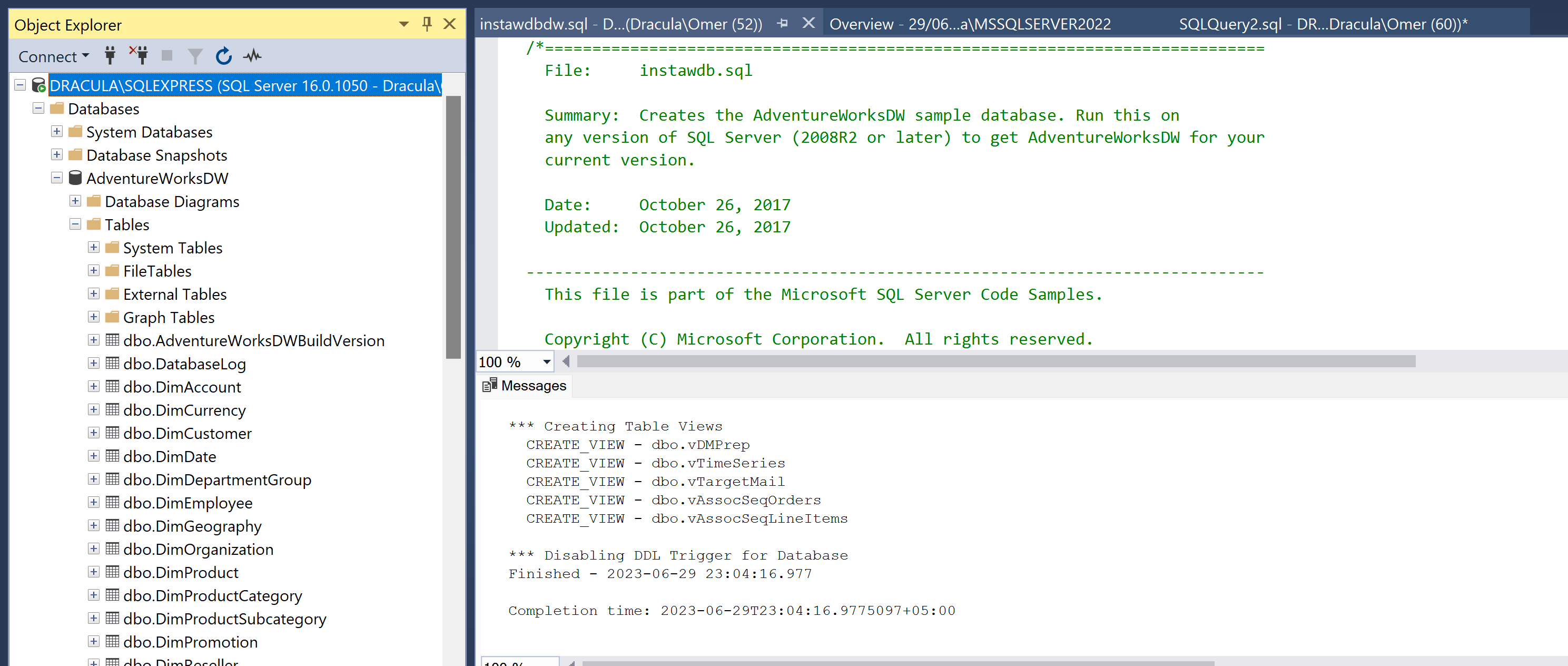
Question 1: Download and import AdventureWorksDW2012 in the MS SQL Server. If you want you can use this script to create AdventureWorksDW2012.



Question 2: The AdventureWorksDW2012 Internet Sales Schema Create database diagram using MS Server Management Studio. Use the following Tables to create diagram.

DimCustomer

DimDate

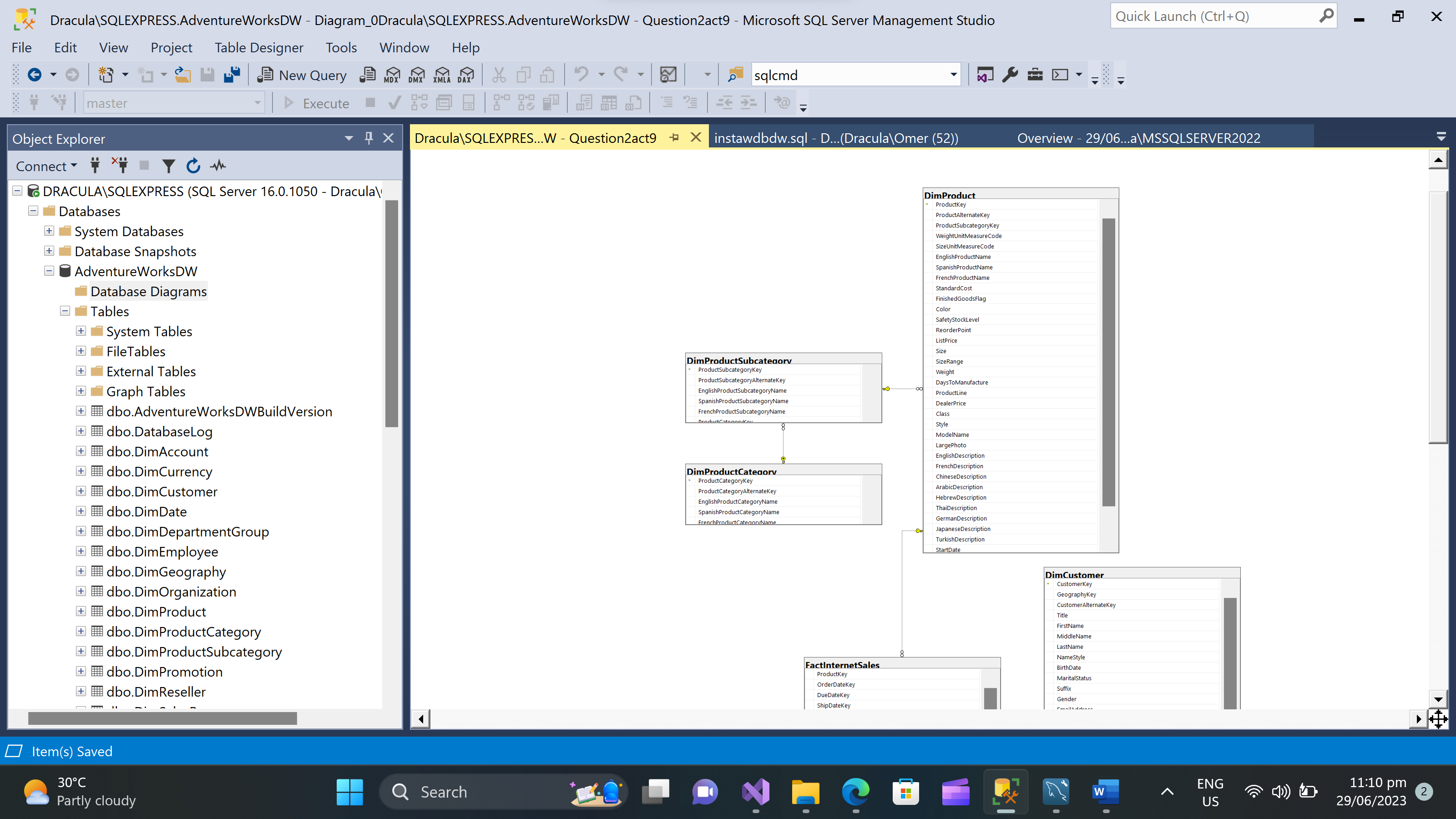
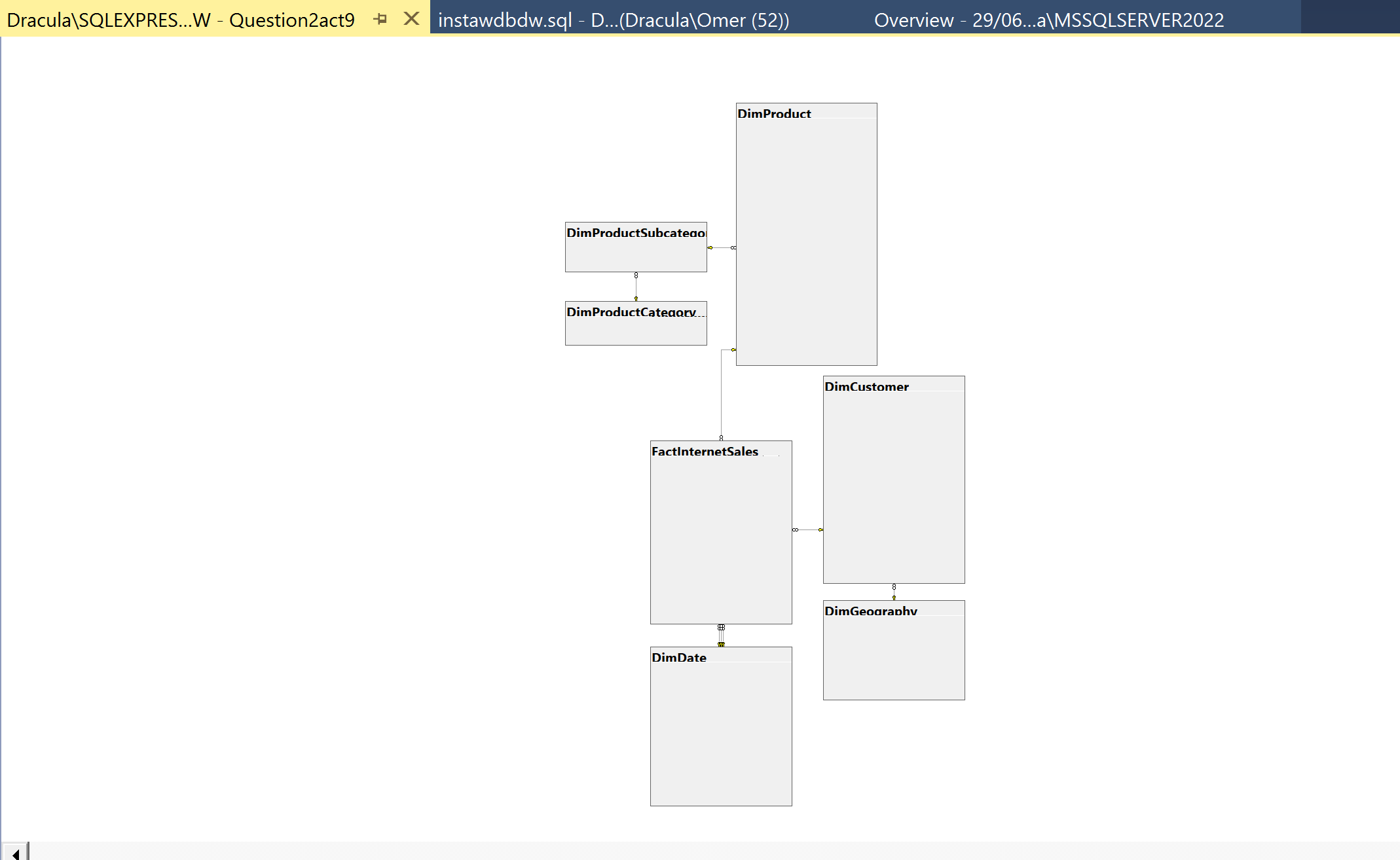
DimGeography

DimProduct

DimProductCategory

DimProductSubcategory

FactInternetSales



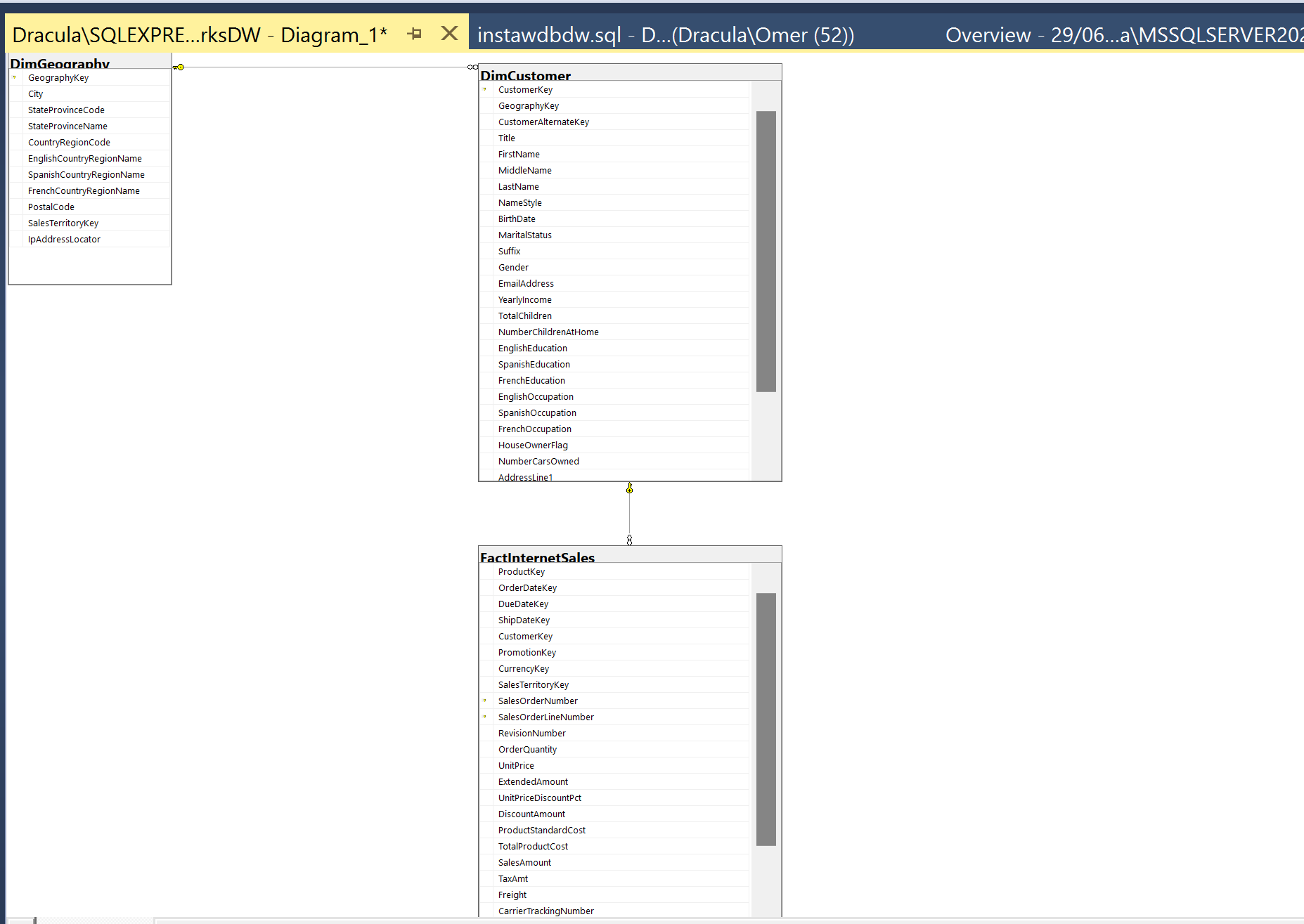
Question 3: Review the AdventureWorksDW2012 schema and answer the following facts:

a. The DimDate dimension has \_\_\_\_\_5\_\_\_\_ additional lookup tables associated with it and therefore uses the Snowflake schema.

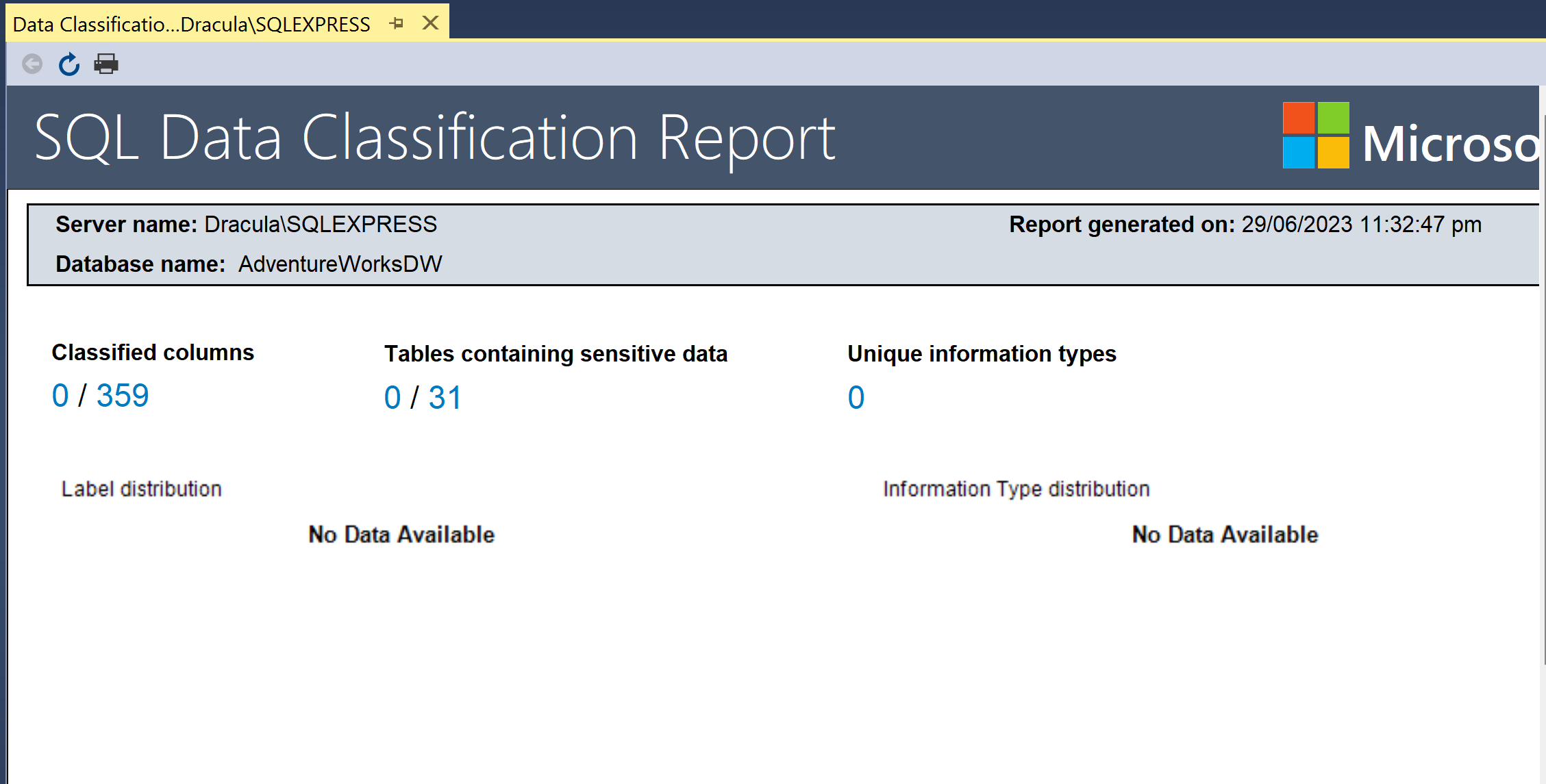
b. The DimProduct table is normalized it uses the DimProductSubcategory

lookup table, which further uses the DimProductCategory lookup table.

c. What tables are connected to DimCustomer dimension. Draw the diagram.



d. List the name of tables that stores lineage information.



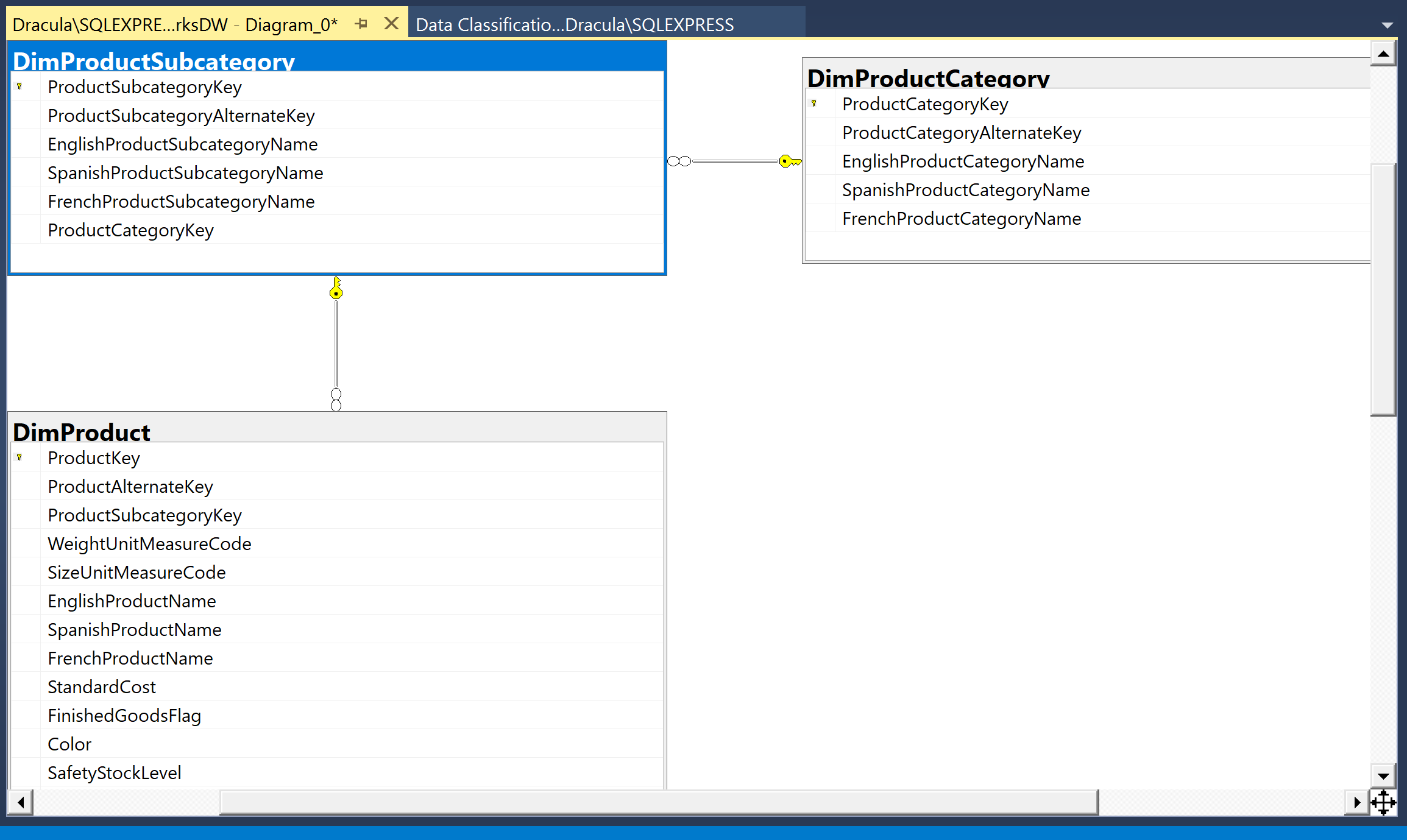
The process of implementing**Data Lineage** in an**ETL** process is very simple. We just have to identify a set of records for the first time, as it enters the**ETL** pipeline.(Bing)

Question 4: The AdventureWorksDW2012 Internet Sales Schema Create database diagram using MS Server Management Studio. Use the following Tables to create diagram.

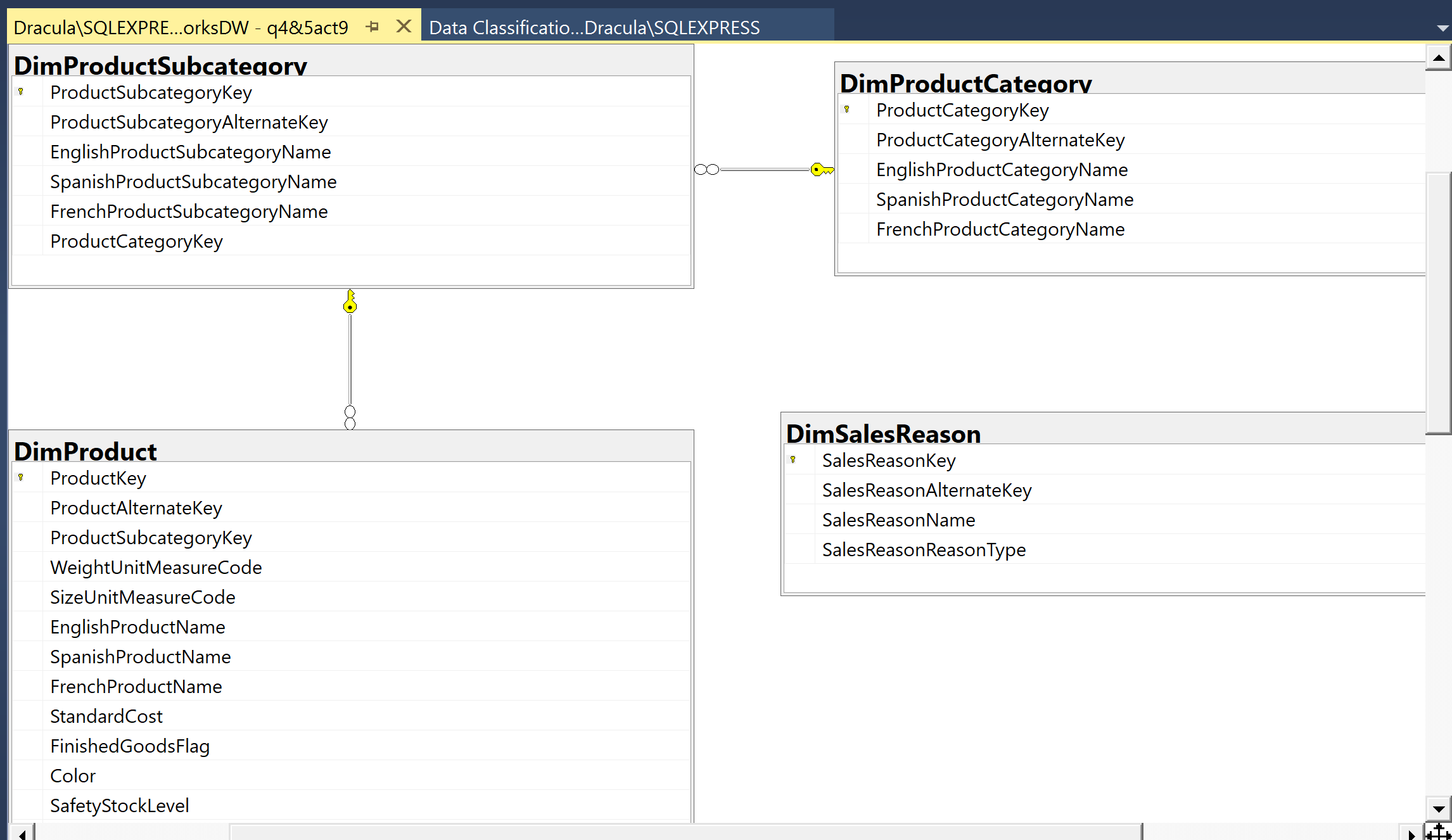
DimProduct

DimProductCategory

DimProductSubcategory



Question 5: Draw the database diagram using MS Server Management Studio by adding the DimSalesReason table to the previously created diagram in last question.



Question 6: Using the previously created diagram, answer the following

a. The DimProduct dimension has a natural hierarchy: ProductCategory ➝ ProductSubcategory ➝ Product.

b. The DimProduct dimension has many additional attributes that are useful for analysis but that are not a part of any natural hierarchy. For example, Color and Size are such attributes.

c. Some columns in the DimProduct dimension, such as the LargePhoto and Description columns, are non-additive (respect to dimension column type).

d. DimSalesReason uses a normalized schema.

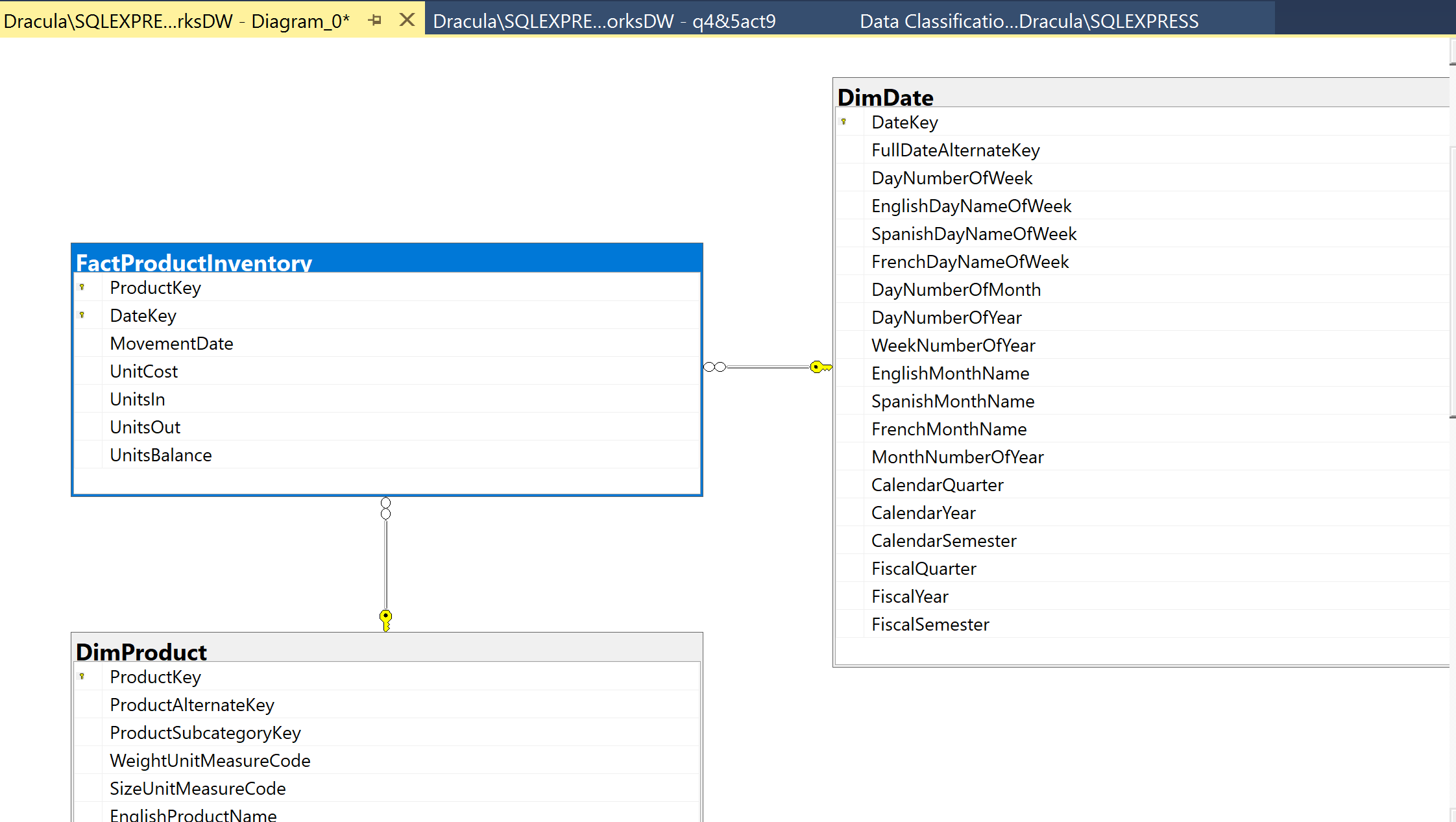
e. In the DimSalesReason dimension, it seems that there is a natural hierarchy: SalesReasonType ➝ SalesReason

Question 7: The AdventureWorksDW2012 Internet Sales Schema Create database diagram using MS Server Management Studio. Use the following Tables to create diagram.

DimProduct

DimDate

FactProductInventory



Question 8: Analyze Fact Table Columns Learn more details about the fact table in the schema you created in the previous question. Note that you have to conclude these details from the names of the measure columns; in a real-life project, you should check the content of the columns as well.

1. Knowing how an inventory works, you can conclude that the UnitsIn and UnitsOut are additive measures. Using the SUM aggregate function for these two columns is reasonable for aggregations over any dimension.

Aggregating these measures would provide meaningful results when analyzing inventory movements.

1. The UnitCost measure is a non-additive measure. Summing it over any dimension does not make sense.

Because it represents the cost per unit and is not meant to be aggregated. Instead, it is typically used for calculating the total cost or average cost at a specific level of granularity.

1. The UnitsBalance measure is a semi-additive measure. You can use the SUM aggregate function over any dimension but time.

Aggregating the UnitsBalance measure over time would not provide meaningful results since it represents the balance or quantity of inventory at a specific point in time.

Question 9: Case Scenario 1: A Quick POC Project You are hired to implement a quick proof of concept (POC) data warehousing project. You have to prepare the schema for sales data. Your customer’s subject matter experts (SME) would like to analyze sales data over customers, products, and time. Before creating a DW and tables, you need to make a couple of decisions and answer a couple of questions:

1. What kind of schema would you use?

A suitable schema would be a Star schema. The Star schema is simple, denormalized, and easy to understand, making it ideal for quick prototyping and analysis purposes.

1. What would the dimensions of your schema be?

Customer dimension: This dimension would contain customer-related attributes such as customer ID, name, address, demographics, etc.

Product dimension: This dimension would include product-related attributes such as product ID, name, category, price, etc.

Time dimension: This dimension would capture time-related attributes such as date, month, quarter, year, etc.

1. Do you expect additive measures only?

When analyzing sales data, it is important to consider both additive and non-additive measures. Additive measures, like sales amount and quantity sold, can be summed across dimensions, while non-additive measures, such as average selling price and profit margin, cannot be easily aggregated. By incorporating both types of measures in the data warehousing project, a comprehensive understanding of the sales data can be achieved, allowing for a wide range of analytical possibilities, including dimension-based aggregations and calculations of average values.

Question 10: Case Scenario 2: Extending the POC Project After you implemented the POC sales data warehouse in Case Scenario 1, your customer was very satisfied. In fact, the business would like to extend the project to a real, long-term data warehouse. However, when interviewing analysts, you also discovered some points of dissatisfaction. Interviews Here’s a list of company personnel who expressed some dissatisfaction during their interviews, along with their statements:

• Sales SME “I don’t see correct aggregates over regions for historical data.”

• DBA Who Creates Reports “My queries are still complicated, with many joins.” You need to solve these issues by answering the following questions

1. How would you address the Sales SME issue?

To address the Sales SME's concern regarding incorrect aggregates over regions for historical data, modifications to the schema and data model may be needed. This can involve reviewing and adjusting the dimension hierarchy, ensuring proper association of historical data, and refining the ETL process for data transformation and loading.

1. What kind of schema would you implement for a long-term DW?

For a long-term data warehouse project, a Snowflake schema is suitable due to its normalized structure. It reduces data redundancy and improves integrity, making it ideal for complex and scalable projects that involve large volumes of data over an extended period.

1. How would you address the DBA’s issue?

To tackle the DBA's challenge of complicated queries with multiple joins, options include denormalization of dimensions, creating summary tables, implementing materialized views for commonly used queries, and optimizing query execution through indexes, statistics, and query plans.

By implementing these solutions, the concerns raised by the Sales SME and DBA can be addressed, providing accurate aggregates over regions and simplifying queries for improved performance.