Assignment

1. Operational database

The operational database is provided in the AdventureWorks2014.bak and AdventureWorks2019.bak files. Depending on the version of your sql server, restore the corresponding file (AdventureWorks2019.bak for SQL Server 2019 and AdventureWorks2014.bak for previous versions of SQL Server) for having the operational database.

2. Context

Suppose that you are an employee of a large company that sells bicycles and cycle accessories. This company has sales in Canada, the United States and in various Spanish-speaking countries in South America such as Mexico. The diagram the operational base is too large to print on a page A4, you could create a database diagram in your SSMS to view the operational base. The operational base is divided into several schemas (Sales, Person,Production ...), each corresponding to a specific operational need. Note that:

- The SalesOrderHeader table contains general information about customer orders; The CurrencyRateID = NULL corresponding to order paid in USD.
- The SalesOrderDetail table contains the details of the products sold.
- The SalesTerritory table contains geographic information about customers. It also serves to identify the geographical areas covered by the company (although this is not what interests us here).
- The Customer table contains the customer information. The customers can be of two types: individual customers (having PersonID not null) or store (having StoreID not null). Note that you have also a list of old customers provided in the file Customers.xls. There could have some customers that appear both in the Customer table and in the excel file.

Business Requirement

The marketing department wants to increase the volume of online sales (SalesOrderHeader.OnlineOrderFlag = 1) for individual customers (customers having PersonID not null). For this, it is necessary to better describe and understand the way in which these sales work.

Your goal is to build a datamart regarding sales made to individual customers (individual). We do not consider here sales made for professional customers (store), moreover these stores get their supplies directly from a commercial, as you can see in the diagram operational. Your analysis should focus on **online** orders that have already been delivered to **individual** customers.

Your analysis must relate to each product: its characteristics (category, subcategory, model ...), their cost, etc ... For each product, the manager needs to analyze by name, color, standardCost, ListPrice, Size,

SizeUnitMeasure, Weight, WeightUnitMeasure, ProductLine, Class, Style, Category, Subcategory, Model. Note that the decision makers do not want to analyze the sales for each precise price, cost or weight but they want to analyze the sales according to each price range (0-100, 100-200, 200-300, …, 900-1000, 1000-2000, 2000-3000, 3000-4000, >= 4000), to each cost range (0-100, 100-200, 200-300, …, 900-1000, 1000-2000, 2000-3000, 3000-4000, >= 4000) and to each weight range (0-50, 50-100, 100-150, 150-200, 200-250, 250-300, >= 300).

We are interested also in the geographic area where the sale took place (SalesTerritoryRegion (Name attribute in SalesTerritory table), SalesTerritoryCountry (note that in the data warehouse, we are interested in the country region name, not the CountryRegionCode), and SalesTerritoryGroup), as well as the promotions (Special offer) used by customers during their orders. Promotions are in the form of percentages of reduction. The promotion can be some types, it specifies a minimum and maximum number of products purchased (MinQty and MaxQty), and has time limits (StartDate and EndDate).

Your analysis should also look at the dates the client has ordered products (OrderDate), on the dates that these orders should have been delivered (i.e. the DueDate that was communicated to the customer at his request), and the dates on which the order was delivered to them (ShipDate). Indeed, long waiting times between the order and the delivery or if the customer has to wait much longer than what was announced when ordering, this can have a very negative impact on future sales for this client, which could prefer to go to a competitor's store in the future rather than continue to place his internet orders in our society.

Customers (title, gender, etc.) are of particular interest to you. In this project, you will also create customer profiles to help members of the marketing department in the implementation of targeted marketing. Decision makers of the marketing department have explicitly asked to **keep all the information on their clients**, namely Title, FirstName, MiddleName, LastName, Gender, BirthDate, MaritalStatus, EmailAddress, YearlyIncome, TotalChildren, NumberChildrenAtHome, Education, Occupation, HomeOwnerFlag, NumberCarsOwned, AddressLine1, AddressLine2, City, PostalCode, State, Country, PhoneNumber.

Your analysis should also follow the currency in which internet sales are made (Euro, US $ ...) and the amount paid in that currency, but without going into the detail of exchange rates. Note that in the SaleOrderHeaders table, CurrencyRateID = NULL corresponding to order paid in USD.

Of course, the margin will be studied closely, and therefore the costs of the products sold and the amount of tax will have to be taken into account. Note that, since this company has sales in Canada, the United States and various countries Hispanics in South America such as Mexico, VAT rates may change in function of the sale. The price of delivery (freight) will also be taken into account because it will have a direct impact on the margin, obviously.

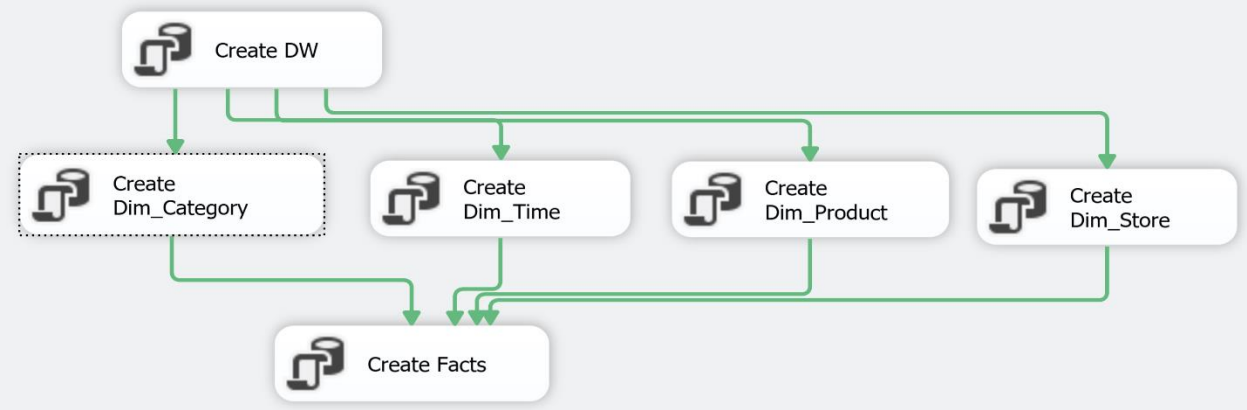
Your work:

1. **For the lab 4:**
    a. **Propose an information package for the data mart that you must create.**
    b. **Distinguish the categories/hierarchies in each dimension.**
    c. **Give the level of details of the data that should be stored in your data mart.**
    d. **Using star schema to construct a Multi Dimension Data Model.**

e. **Enrich your star schema as follows: if an attribute of a dimensional model table comes from an attribute of the operational database, specify in parentheses the name of this attribute source. Otherwise, bring up all the attributes of the database from which this new attribute can be deduced.**

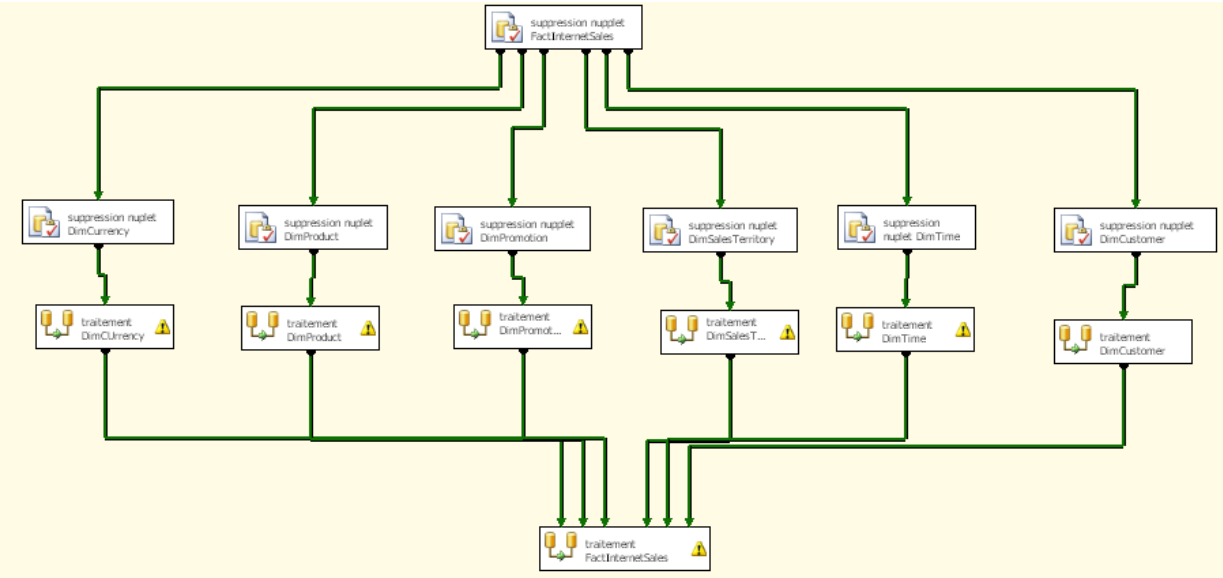
2. **For the lab 5:**

   Use SQL Server Integration Services to create a new Integration Services project named Bicycles_ETL.sln. In this project, you will create a new Package named Create_DW.dtsx. In this package, you will create the multi dimension database (data warehouse) and the corresponding tables. Specifically, in this package, you have a "control flow" in which you can add different tasks to this control flow. These tasks can be organized sequentially or parallel, as shown in the following figure. This figure shows only an example, it is not corresponding to the control flow of your project. In fact, you must create a "SQL query task" for creating the database or for creating each table. And for each task, you could put your Transact-SQL code in the "SQL-query" field corresponding to the task you want.



3. **For the lab 6:**

   In the project Bicycles_ETL.sln, create a new package named Populate_Decisional_DB.dtsx. In this package, you will create different "data flow task" in order to feed (ETL process) your data warehouse from the operational database as well as the customers data from the excel file. An example of the ETL process is given in the following figure. Note that information of Customers exists both in the operational database as well as in the Customer.xls file. So, you must combine information from these two sources.



4. **In the project Bicycles_ETL.sln, create a new package named Incremental_Load_DB.dtsx for doing the incremental load (for example, each night) for the DimCustomer table.**
5. **Data analyst: Use SQL server Analysis Services to create an Analysis Service project named "Bicycles_analysis", create a cube including all dimensions and facts and then try to navigate in the cube for analyzing the data and create analyst table as you wish. Try to do an analysis as you wish by:**
    a. **Navigating in the cube that you created.**
    b. **Creating SQL query in the operational database.**
    c. **Creating SQL query in the data mart.**

    An example is to analyze the total quantity sold by category, by customer gender and by year.

6. **Use SQL Server Reporting Service to create a "reporting service" project by using your data warehouse as data source and then create a report corresponding to the analysis that you did in the previous question and extract the report in an excel file. You must also public an online version of your report in your local server http://localhost/ReportServer.**