

Hand in 1

AdventureWorks is a large multinational manufacturing company, producing and selling metal and composite bicycles to the international market. Elisabeth, the sales manager, has requested a data warehouse to be designed to analyze AdventureWorks' profitability throughout the course of the year in order to improve future sales performance, especially in the off-season and assist the business transition towards the "direct to customer" model.

Dimensional model

Business process: Sales profit

Elisabeth requested to focus on the profitability of AdventureWorks' sales.

Grain :

Order line item - A sale can be composed of many items and the lowest granularity is at the order item line. Order item lines are entered into the SalesOrderDetails table in the AdventureWorks2019 source database.

Dimensions :

The dimensions are selected as :

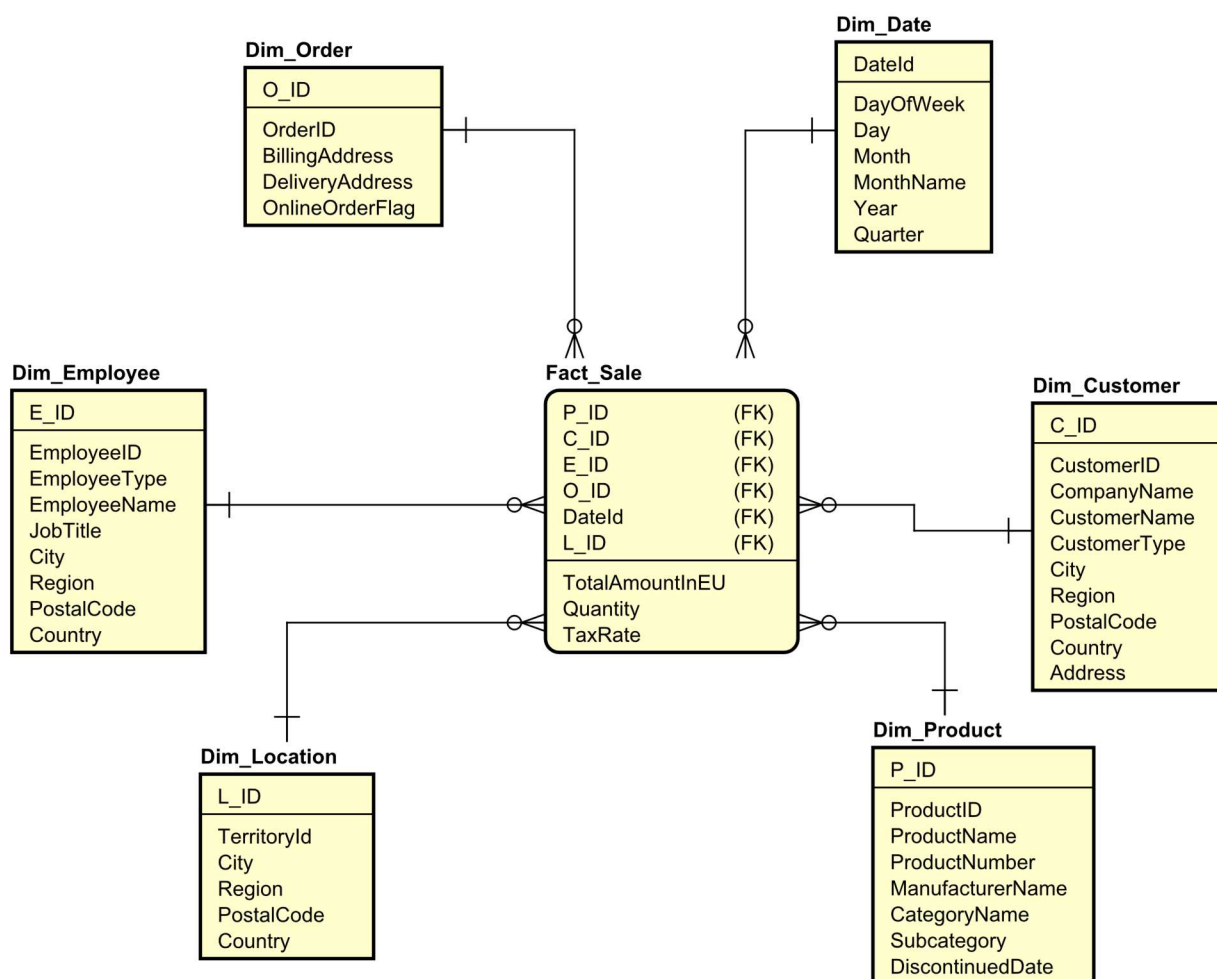
- Customer
- Employee
- Location
- Date
- Product
- Order

These dimensions can be derived from the case description because Elisabeth asks for analysis on **sales** for different types of **customers**: individual sales and store sales. Given that there are 2 types of **customers**, the **orders** they place are divided into wholesale and online **orders**. The **employee** and **product** dimensions were added as Elisabeth wants to be able to look at the sales performance of different sales **personnel** with respect to both the total amount and number of items sold across **product** categories. She also asks for data to broaden its market share by targeting the sales to their best customers, so we need a geographical **location** dimension as well. Furthermore, Elisabeth wants to be able to come up with some strategies to improve sales in the off-season months therefore we need a **date** dimension.

Facts :

Elisabeth is asking for analyses primarily of profitability and sales. The customer wants to see the difference with respect to the number of items a wholesale customer orders compared to a private consumer. Therefore, we need metrics that will allow us to do these analyses.

From Sales Order, we can see the sales performance of different sales personnel with respect to both the **total amount** and **number of items** sold across product categories. We can see that the measurable events in the source database are OrderQuantity, LineTotal amount and TaxRate, therefore they were defined as facts. Since Adventure Works is a multinational company, the sales are done in different countries, therefore the currency is different. In order to make it easier for Elisabeth to see the profit of the sales and compare it, the total amount is converted in euros according to the currency in the country. Information for the conversion rate can be found in the CurrencyRate table in the source database.



ER Diagram showing the star schema

Dimensions and Attributes:

For Customer, we have included the following attributes: **C_ID, CustomerID, CompanyName, CustomerName, CustomerType, City, Region, PostalCode, Country, Address.**

C_ID is the primary surrogate key for the table. CustomerID is the primary key from the source table and is the business key for the Customer. CustomerName displays the name of the customer., while CustomerType indicates whether the customer is an individual or a business. The customer's location is indicated by the City, Region, PostalCode, Country and Address.

For Employee, we have included the following attributes: **E_ID, EmployeeID, EmployeeType, EmployeeName, JobTitle, City, Region, PostalCode, Country**
E_ID is the primary surrogate key for the table. EmployeeID is the primary key from the source table and is the business key for the Employee. EmployeeName is going to be used to display the names of the employees for easy distinguishing. JobTitle shows if the employee is a Sales Representative or something else. City, Region, PostalCode, Country and Address show the location of the employee.

For Product, we have included the following attributes: **P_ID, ProductName, ManufacturerName, CategoryName, SubcategoryName, DiscontinuedDate**
P_ID is the surrogate and primary key for the table. For a data warehouse, we need a PK to distinguish between different versions of the same product in order to allow change tracking (keep track of history)

ProductID is the business key for products.

The name fields are included so that after Elizabeth finishes her visualizations, she has a user-friendly summary of the data with the proper labels. E.g: CategoryName because she wants to be able to look at the sales performance of different sales personnel sold across product **categories**.

DiscontinuedDate to be able to see when the product is no longer available for customers.

For Order, we have included the following attributes: **O_ID, BillingAddress, DeliveryAddress, OnlineOrderFlag**

O_ID is the primary surrogate key for the table. BillingAddress is the residential address of the customer who made the order. The DeliveryAddress is the address to which the order should be delivered. OnlineOrderFlag signifies if the order was made on the website or in a store.

For Date, we have included the following attributes: **Dateld, DayOfWeek, Day, Month, MonthName, Year, Quarter**

Dateld is the primary surrogate key for the table.

DayOfWeek, Day, Month, MonthName, Year are included so that Elisabeth can easily visualize the data.

The Quarter is included so that Elisabeth can see the sales performed in the different months of the year, especially the winter months as she is concerned that the business is not making enough money during this period.

For Location, we have included the following attributes: **L_ID, TerritoryId, City, Region, PostalCode, Country**

L_ID is the primary surrogate key for the table.

TerritoryId because Elizabeth is interested in comparing AdventureWorks different geographical locations in terms of total sales in order to track which markets are performing and which are not.

City, Region, PostalCode and country in order for her to visualize the data.

Data types:

For data types for all string types, we have chosen nvarchar as it is the recommended data type to use when the size of the entries vary considerably [1]. However, for CustomerType we have chosen nchar(2) as it is only 2 characters in the source database. For IDs, we have chosen int, as they are most likely integer values.

Fact table design:

For the fact table, we have included all the primary keys to map to dimensions.

We have included the total amountInEU converted based on location, quantity and tax rate in order to be able to model “ the sales performance of different sales personnel” as requested by Elisabeth.

Install guide:

The data warehouse is installed by first running 1-CreateAdventureWorksDW1.sql which creates the database for the data warehouse and the tables.

Source target mappings

Source Target Mappings are available in the appendix (excel sheet “Source-target-mappings”). The example is seen below:

Source System: Adventure Works Cycles Database						
Stage						
Source			Target			
Table	Field	Datatype	Table	Field	Datatype	
			Dim_Product	P_ID	int	<i>created in source sytem</i>
Product	ProductID	int	Dim_Product	Product_ID	int	
Product	Name	nvarchar(50)	Dim_Product	ProductName	nvarchar(50)	
Product	ProductNumber	nvarchar(25)	Dim_Product	ProductNumber	nvarchar(25)	
			Dim_Product	ManufacturerName	nvarchar(50)	<i>created in source sytem</i>
ProductCategory	Name	nvarchar(50)	Dim_Product	CategoryName	nvarchar(50)	
ProductSubcategory	Name	nvarchar(50)	Dim_Product	SubCategory	nvarchar(50)	
Product	DiscontinuedDate	datetime	Dim_Product	DiscontinuedDate	datetime	

For Employee/Customer name, the target column is EmployeeName/CustomerName and the source fields are a concatenation of First Name, Middle Name and Last Name from the source system.

[1] Source: <https://docs.microsoft.com/en-us/sql/t-sql/data-types/nchar-and-nvarchar-transact-sql?view=sql-server-ver15>