Hand in 1

***Topic:***

Initial analysis, design and implementation of the dimensional model required for the AdventureWorks case. (Inspiration, reference case – NorthWind)

***Given case:***

***Sales reporting****. Senior management would like to be able to track sales by customer and product, with the goal of establishing which products are the top sellers, which customers are the top buyers, and sales per month for top 3 selling products.*

***Tasks:***

* *Identify the grain*
* *Identify the dimensions*
* *Identify the facts*
* *Do an initial design of the Data Warehouse/Data Mart.*

1. *The 4-step design process:*

Initial: Analysis

For the initial identification of the fact grain, a similar grain was identified to that of the example from the NorthWind case as the definition of the business process is nearly the same. Moreover, analysis of the source database was made prior to the modelling phase as to view the state of the metadata.

The identification of the dimensions was with consideration to the business requirements that were analysed, and keywords noted.

*Sales reporting. Senior management would like to be able to track sales by customer and product, with the goal of establishing which products are the top sellers, which customers are the top buyers, and sales per month for top 3 selling products.*

***The business requirements needed for the management’s analysis:***

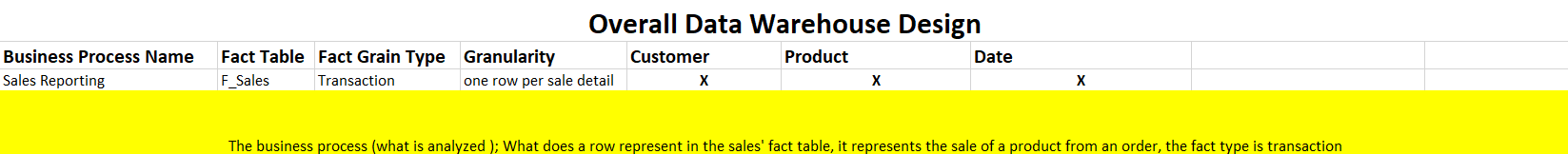
* Sales by customer
* Sales by product
* Sales per month

Step 1: Business process identification

Business process: ***Sales reporting***

Step 2: Declare the grain (for the fact table)

Fact table Grain (Fact Row definition)



Step 3: Identify the dimensions (contextual information)

*The customer, product, and date dimensions to support in the realization of the sales fact (satisfying the business requirements).*

* D\_Customer (Customer dimension)

The customer dimension has C\_ID as a surrogate primary key. The other attributes are, Name (concatenated attribute from FirstName, MiddleName, LastName to avoid unnecessary details and null values as several records have MiddleName with the null as a value) and City, all of which have as a purpose identification of adherent descriptive customer information, the City [[1]](#footnote-1) attribute allows grouping of customer records.

* D\_Product (Product dimension)

The product dimension has P\_ID as a surrogate primary key. The other attributes, ProductID, ProductName, ProductNumber, ProductCategory, ProductSubCategory are being used for descriptions of products that are part of a sale and for later filtering of records.

* D\_Date (Date dimension)

The date dimension has D\_ID as a surrogate primary key. The other attributes are CalendarDate, WeekDayName, MonthName and WeekDayNumber. Except for the primary key, all the attributes aim at filtering of sale dates. Essential is the MonthName

attribute as it allows identification of specific months in conjunction with sales as per the business requirements.

Step 4: Identify the facts (metrics)

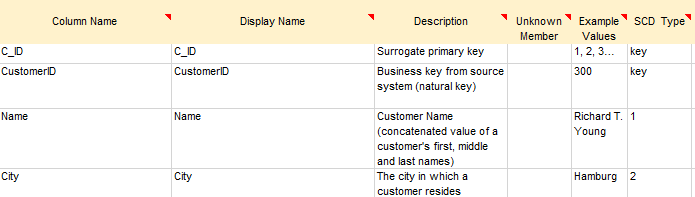
**Facts:**

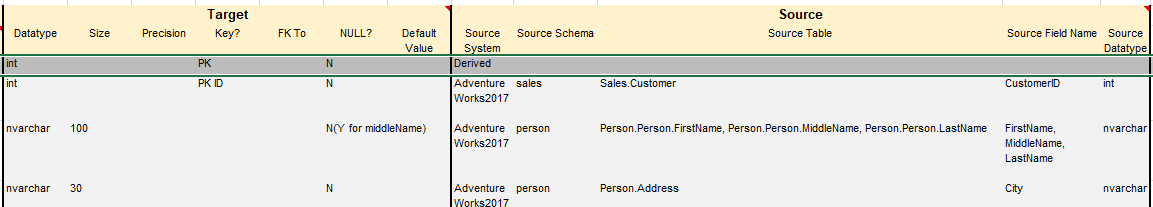
* LineTotal
* OrderQty
* F\_Sales (Sales fact)

Table created to satisfy the requirement of tracking sales by customer and product, with the goal of establishing which products are the top sellers, which customers are the top buyers, and sales per month for the top 3 selling products. The table has a composite primary key (P\_ID, C\_ID, D\_ID) of which all consisting attributes are foreign keys to their respective dimensions. F\_Sales has the goal of encapsulating metrical data such as /LineTotal [derived attribute from ((UnitPrice – UnitPriceDiscount) \* Quantity) ] together with the ordered quantity of the product within the sale – OrderQty[[2]](#footnote-2).

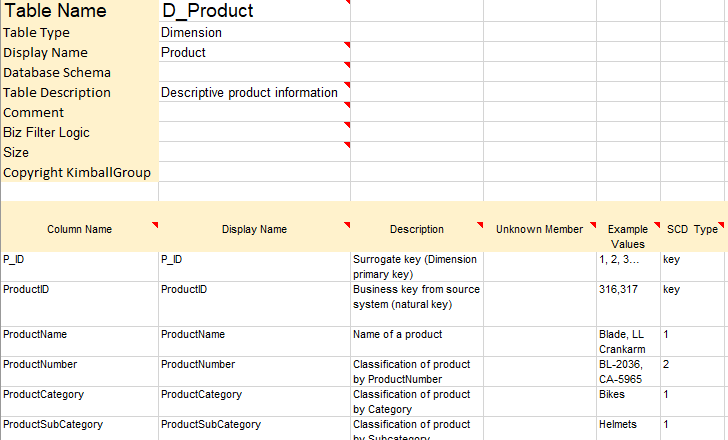
1. Design documentation (blueprint) AdventureWorks2017\_DataWarehouse

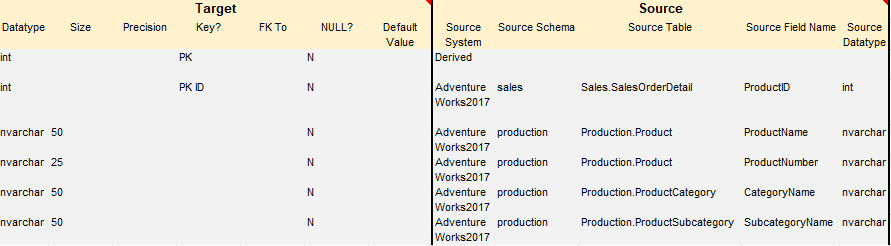
* D\_Customer (The Customer dimension)



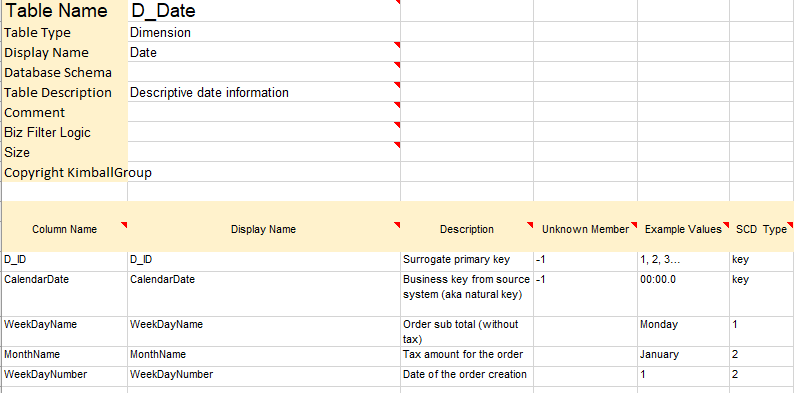


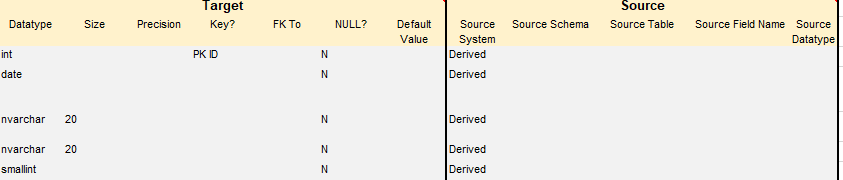
* D\_Product (The Product dimension)



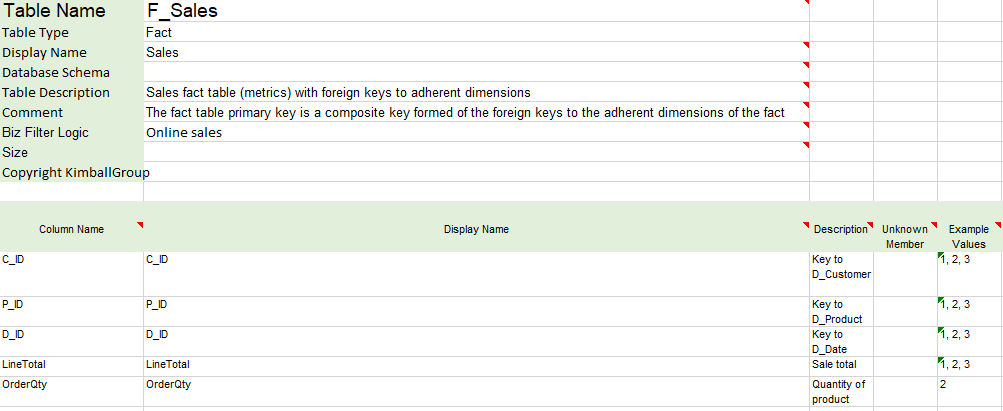


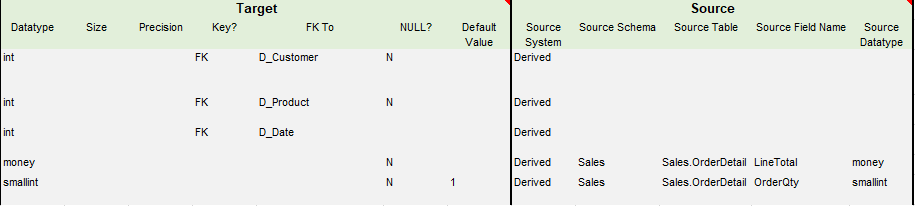
* D\_Date (The Date dimension)



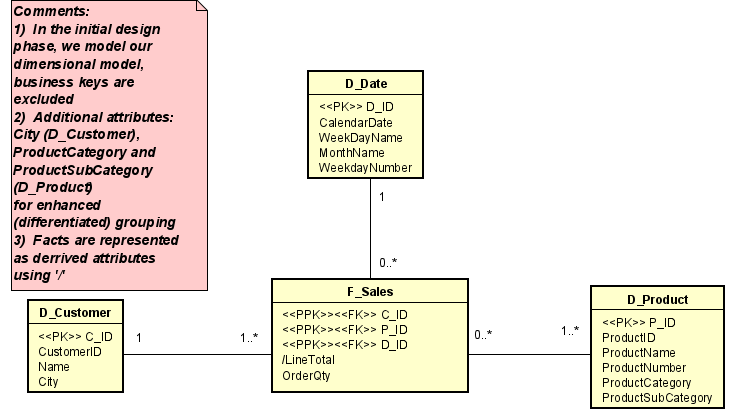


* F\_Sales (The Sales fact)





1. ERD (Star schema) AdventureWorks2017\_DataWarehouse



1. SQL Code (Implementation)
2. Customer dimension implementation (initial – v 1.0)

CREATE TABLE [D\_Customer](

C\_ID [int] IDENTITY(1,1) NOT NULL,

Customer\_ID [int] NOT NULL,

Name [nvarchar](100) NOT NULL,

City [nvarchar](30) NOT NULL,

PRIMARY KEY(C\_ID)

);

1. Product dimension implementation (initial – v 1.0)

CREATE TABLE [D\_Product](

P\_ID [int] IDENTITY(1,1) NOT NULL,

ProductID [int] NOT NULL,

ProductName [nvarchar](50) NOT NULL,

ProductNumber [nvarchar] (25) NOT NULL,

ProductCategory[nvarchar] (50) NOT NULL,

ProductSubCategory[nvarchar] (50) NOT NULL,

PRIMARY KEY(P\_ID)

);

1. Date dimension implementation (initial – v 1.0)

CREATE TABLE [D\_Date](

D\_ID [int] IDENTITY(1,1) NOT NULL,

CalendarDate [date],

WeekDayName [nvarchar](20),

MonthName [nvarchar](20),

WeekDayNumber [smallint],

PRIMARY KEY(D\_ID)

);

1. Sales fact implementation (initial – v 1.0)

CREATE TABLE [F\_Sales](

C\_ID [int] NOT NULL,

P\_ID [int] NOT NULL,

D\_ID [int] NOT NULL,

LineTotal [money] NOT NULL,

OrderQty [smallint] NOT NULL,

CONSTRAINT FK\_F\_Sales\_0 FOREIGN KEY (C\_ID)

REFERENCES D\_Customer(C\_ID),

CONSTRAINT FK\_F\_Sales\_1 FOREIGN KEY (P\_ID)

REFERENCES D\_Product(P\_ID),

CONSTRAINT FK\_F\_Sales\_2 FOREIGN KEY (D\_ID)

REFERENCES D\_Date(D\_ID),

CONSTRAINT PK\_F\_Sales PRIMARY KEY (C\_ID, P\_ID, D\_ID)

);

1. As customers are present within multiple areas of the world [↑](#footnote-ref-1)
2. Used for identification of top selling products (quantity-wise) [↑](#footnote-ref-2)