# Data Warehouse and BI for North Wind Company

Contents

[Data Warehouse and BI for North Wind Company 1](#_Toc507951257)

[Business Drivers 3](#_Toc507951258)

[Data Modelling 4](#_Toc507951259)

[Four Step Process 4](#_Toc507951260)

[Dimensional Model Approach 5](#_Toc507951261)

[Data Warehouse Tools 7](#_Toc507951262)

[Microsoft Solution 7](#_Toc507951263)

[Reporting and Analysis 8](#_Toc507951264)

[Employees 8](#_Toc507951265)

## Business Drivers

Our business drivers behind the building of this data warehouse are to summarise sales figures across the business and identify sales trends. The main questions that are driving the structure and content of the data warehouse are:

* Who is selling well? (Employees)
* What Products are selling well? (Products)
* Where are we selling the most? (Customers)
* When are we selling well? (Date)

The subject area is sales trends, with the data made available to us it is possible to build a data warehouse for the sales team in order to analyse sales trends.

The objective of the data warehouse to bring together all the information that is pertinent to sales, these will answer the questions above and allow the sales leaders to identify sales patterns and make decisions on areas to target or improve.

## Data Modelling

### Four Step Process

The following model was created using design principles outlined in **Kimball (insert reference here)**. Kimball outlines that there are four steps to consider when developing a dimensional model:

1. Choose the business process
2. Declare the grain
3. Identify the dimensions
4. Identify the fact

#### Choose the Business Process

The business process for the Northwind Trading company is to sell and ship products worldwide. A sales team takes orders for companies based on products and quantities this defines the gross sales figure, however discounts can be applied to orders so we have a figure for gross profits and gross profits after discount. Each order that is taken in by a member of the sales team represents the sale of a single product or a number of products and the number of the products sold.

#### Declare the Grain

Each order line that is processed represents a product sold, the employee that sold it, the customer it was sold to, the date it was sold, what the gross sale value on the order was and the gross sale value on the order after discount.

#### Identify the Dimensions

Our dimensions fall in line with our key questions outline above. The dimensions for our data warehouse are the below:

* Date
* Customers
* Products
* Employees

These dimensions represent the who, what and when of our fact table. It is worth noting that the Date dimension is a table created to extract more detailed date information from our order dates.

#### Identify the Fact

Our facts collected through each invoicing order captured in the system incudes the products within an order, the gross sale value, their discounted gross sale value, the date the order was made, the customer who placed the order and the employee who created the order. The gross sale value is made up of the unit price of the product sold by it’s quantity. The discounted sale value consists of the gross sale value multiplied by the order discount amount remaining percentage.

The image below represents a more detailed high level overview of our dimensional model, taken from Microsoft Power BI.

### Dimensional Model Approach

#### Star Schema Approach

The following dimensional model approach we have gone for in building this Data Warehouse is the Star Schema approach. Looking again at the high level overview of our dimensional model we can see our fact table linked by associated dimension tables via foreign key relationships. The reason the star schema approach was opted for was due to a star schemas simplicity and symmetry **(Kimball et al. , Data Modelling toolkit)**. Taking a look at our schema, it is very easy to navigate and should be recognisable to the majority of business users.

#### Advantages

The advantage of using a star schema compared to a traditional RDBM is that there is a reduction in the number of tables used, therefore it is easier for a business user to navigate the data available to them. Another advantage of using the star schema approach is that we should see an improvement in performance as there are fewer joins for the database optimisers to process **(Kimball et al. , Data Modelling toolkit)**.

The final advantage we have here for our data model approach is that the dimensional model design is accommodating to change. The reason for this is that as our dimensions are symmetrically equal entry points into our Sales Fact table **(Kimball et al. , Data Modelling toolkit)**. If a new business question arises from the sales team next month our next step then is to simply identify a new dimension and define a single value for that dimension in the sales fact table, assuming it exists.

#### Disadvantages

There are few disadvantages for opting for a dimensional model for our data warehouse. We have identified previously that Star Schemas are simplistic however this means we have had to exclude data captured in the RDBM source for the Northwind Trading Company. Implementing this value into our model would require us to add a snowflake dimension but this adds a complexity to the schema that the business users wanted to avoid.

A recommendation for moving forward may be to rethink how we capture our date dimension, this table captures information relevant to dates from the years 1990 to 2015. This approach is OK to our dataset as it only contains information within the years 1996, 1997 and 1998. However our date dimension is not future proofed should any information be captured after 2015. For now this captures the relevant information around dates that is needed for the years at our disposal.

A screenshot of a cell phone

Description generated with very high confidence

Figure Data Warehouse Schema

## Data Warehouse Tools

### Microsoft Solution

The data warehouse solution we have opted for here is the Microsoft solution, SQL Server 2014. Microsoft is recognised as a leader in the Gartner magic quadrant for Data Management Solutions for Analystics. The Forrester Wave report for Big Data Warehouse also recognises Microsoft as a significant vendor, the report identified Microsoft’s diverse application platform for Enterprise Data Warehouse solutions.

However, Gartner does warn about a vendor lock in with the Microsoft solution particularly in the cloud instance compared with other vendors such as Amazon or Oracle. Microsoft does display a strong market presence and it’s seamless integration with its rich platform suite such as Power BI, makes this a favourable vendor choice for data warehouse architecture.

## Reporting and Analysis

The BI solution we opted to use for our reporting and analysis was Microsofts Power BI solution. As well as Power BI’s rich reporting and analysis tools, this solution fits in well with our Microsoft warehouse stack. Using Power BI we can create dashboards that answer our business process questions:

* Who is selling well? (Employees)
* What Products are selling well? (Products)
* Where are we selling the most? (Customers)
* When are we selling well? (Date)

Using Power BI we created two dashboards, one dashboard which gives us an overview of the breakdown of sales by product, employee, location (city and country) and customer (Figure 7). The other dashboard is a breakdown of sales over time (Figure 8).

### Employees

A picture containing screenshot

Description generated with very high confidenceThe information stored on sales are from July 1996 to May 1998. All sales members were employed in the Northwind Trading Company during this period so this would not affect Sales trends figures over time. It is also worth noting that employees report to either Steve Buchanan or Andrew Fuller but Steve Buchanan reports to Andrew Fuller as well. If we look at the donut chart (figure 1) of employees (including their Employee ID) discounted sales we can see who the highest and lowest earners are. Janet Leverling and Margaret Peacock appear to be the highest earners, while Steve Buchanan and Michael Suyama are the lowest. However these are just the quick insights and we can use our dashboard to drill down deeper into the information. Looking at the stacked column (figure 3) chart we can see that Janet (in red) has the Southern region to herself. Northwind Trading could look to reward Janet for her hard work to date and perhaps look to invest more in the Southern region as that is quite a large intake for one employee. Looking at figure 3 we can see that the Eastern district brings in our largest number of sales, however the lowest earner in this district is Steve Buchanan. The difference between the sales for Steve in the Eastern district and the other employees is quite large. Steve Buchanan has been identified as one of the supervisors, therefore his other duties may be distracting him from his sales duties. Perhaps Steve could be contacted to see if he requires support to increase his sales intake.

Figure Discounted Sale Value from Employee

A screenshot of a cell phone

Description generated with very high confidence

Figure Regional Discount Sale Values with Employees

### Products

Products can be broken down into individual Product items or categories. We can see for individual products that the top three sellers are Cote de Blaye (Beverage), Thrunger Rostbratwurst (Meat/Poultry) and Raclette Courdavault (Figure 4); drilling further into Power BI and show these are all higher value item (Table 1). The largest sellers by categories are Beverages, Dairy Products and Confectionary.

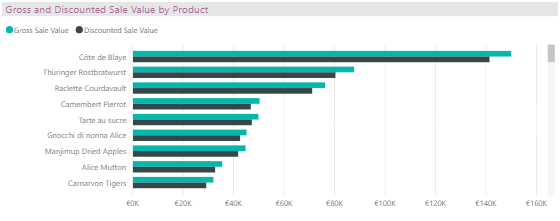


Figure Gross and Discounted Sale Value by Product

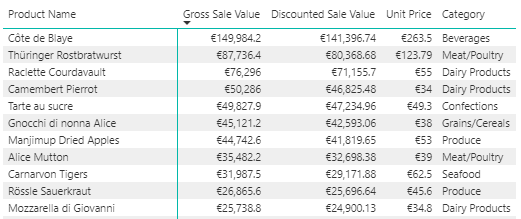


Table Product Sale Values with Unit Price and Category

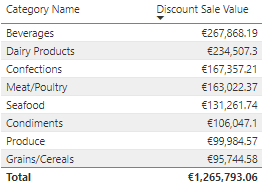


Table Discount Sale Value by Category

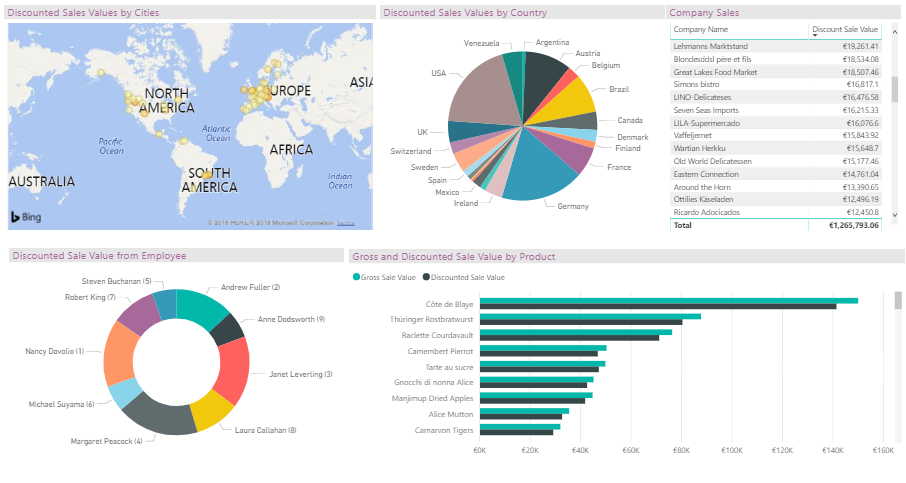


Figure Breakdown of Sales Figures

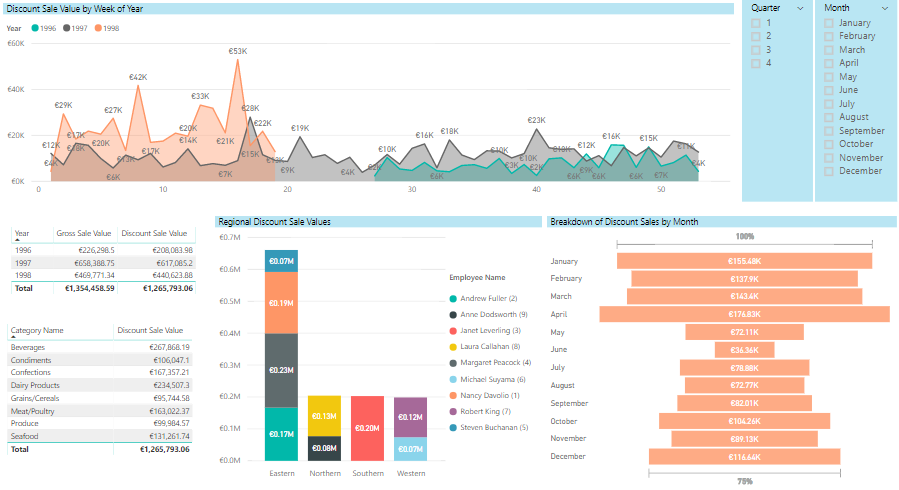


Figure Breakdown of Sales Figures Over Time

### Customers