# Data Warehouse and BI for North Wind Company

Contents

[Data Warehouse and BI for North Wind Company 1](#_Toc507330950)

[Business Drivers 3](#_Toc507330951)

[Data Modelling 4](#_Toc507330952)

[Four Step Process 4](#_Toc507330953)

[Dimensional Model Approach 5](#_Toc507330954)

## 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

### Data Warehouse Tools

#### Microsoft Solution