D: Detailed Dimensional Modeling

# Part 1: Overview

This assignment will introduce the Detailed Dimensional Modeling process. It picks up where the previous process left off with the High Level Dimensional Modeling workbook. In this assignment we will complete the technical designs for our DDS by completing the specifics of the fact and dimension table implementations.

## Goals

Specifically the goals of this assignment are to:

* Understand the goals of the detailed level dimensional modeling process and practice its steps.
* Master completing the detailed modeling workbook as a means to document the technical designs for a Kimball (DDS) data warehouse architecture.
* Create a formal table design for our Northwind DDS, including tables, keys, data types, and indexes so we can create tables and indexes required for our star schemas (ROLAP).
* Identify data sources of our dimensional model so that we can architect and implement the ETL process in a future phase.

## Effort

This assignment can be done individually or with a partner. If you work with a partner, do not simply divide up the work. Collaborate with each other throughout the exercise as if you were working on the same data warehousing team.

## Technical Requirements

To complete this lab you will need the following:

* Access to the course **ist-cs-dw1.ad.syr.edu** SQL Server, and specifically the Northwind Traders database. You should connect to this server before starting the assignment.
* The detailed dimensional modeling Excel Workbook, available in the same place where you got this assignment.
* ***Your completed high-level dimensional modeling workbook from the previous assignment.***
* Microsoft Excel 2007 or higher for editing the worksheets

# Part 2: Walk Through

In this part of the assignment, we will work together to create a detailed level design for the first high level business process: sales reporting. We won’t complete all the dimensions as some of them are conformed and best left to you as the student to complete as an exercise.

## Getting Started

* Connect to your SQL Server using **SQL Server Management Studio** and open the **Northwind** database.
* Open the High-Level-Dimensional-Modeling Excel Workbook, you completed in the previous assignment.
* Open the detail level

## Detailed Design

The **Detail Dimensional Modeling Workbook** we will use comes from the Kimball Consulting Website <http://www.kimballgroup.com>. The Workbook is a detailed technical documentation tool for building a DDS warehouse. It allows you to document the source to target map for your ETL process and contains Excel macros which automate the SQL code generation of fact and dimension tables! ***We will not use these features in this assignment, however we will in the next one!***

**DO THIS:** You should start by **opening the** **Excel Workbook** and reading the section titled **How to use this tool** under the **Home** worksheet, and then read the **ReadMe** tab. This will give you an overview of how to use this workbook. It’s important to do this first as there are many details I will not explain since they are covered in the workbook instructions.

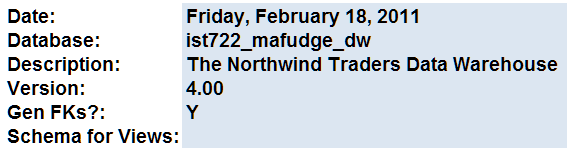
### Getting Started

First let’s setup the workbook.

**DO THIS:** Click on the **Home** tab, and complete the fields as follows:

* Database: **ist722\_*yournetid*\_dw**  (replace *yournetd* with your actual NetID)
* Description: **The Northwind Traders Data Warehouse**
* Gen FK’s?: **Y**
* Schema For Views:  **(leave blank)**

Here’s a screenshot of my completed worksheet.



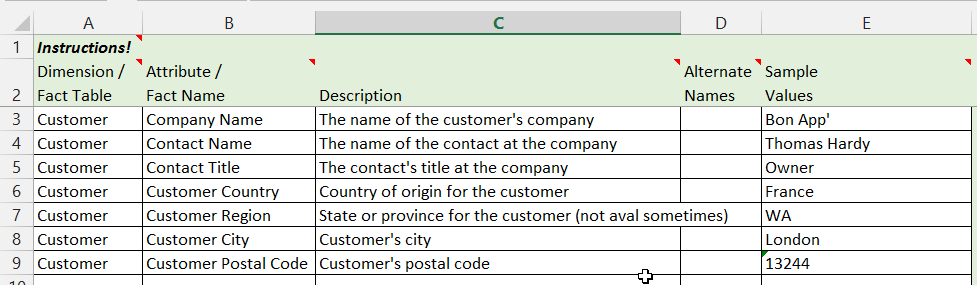
We’re now ready to start our detailed dimensional design.

### Completing the detailed design for the Customer dimension

In this next step, we will complete the detailed dimensional design for the **Customer** dimension. You will need to refer to your **Bus Matrix** and **Attributes & Measures** from the previous assignment to complete this part. The other dimensions will be left as an exercise for you in part 3.

**Important Tip:** The date dimension has been created for you. It shows up under the **DimDate** tab in the workbook.

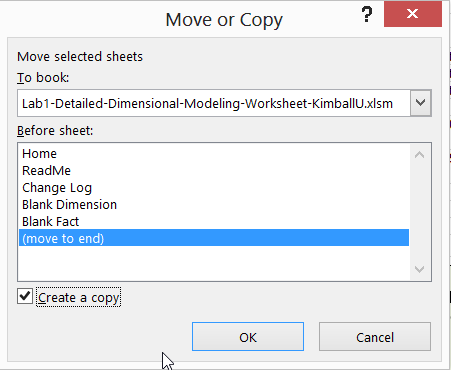
To complete the design we will need to refer to the **Attributes & Measures** from the high level design. A screen shot has been included for reference.



The process you’ll follow to design a dimension or fact table is outlined in **5 Steps:**

1. Create a new dimension (or fact) worksheet in the workbook.
2. Complete the table definition part of the worksheet.
3. Complete the basic column information.
4. Complete the target table information.
5. Complete the source data information.

#### Step 1: Create a new dimension worksheet

Let’s document the details of this dimension.

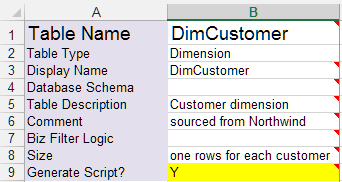
**DO THIS:** Make a copy of the **Blank Dimension** worksheet. ***Right-click*** on it and select ***Move or Copy*** from the menu. When the dialog appears, click ***(move to end)*** and check the ***Create a copy*** checkbox before clicking ***OK***.

You will now have a **Blank Dimension (2)** worksheet. ***Right-click*** on it and select ***rename*** from the menu, then type in **DimCustomer**.

You should now have your first dimension: 

#### Step 2: Complete the table definition

Our next step is to complete the **table definition** for the customers dimension table. Please complete it as follows:



Change to Y

These fields should be completed as necessary

The name of table

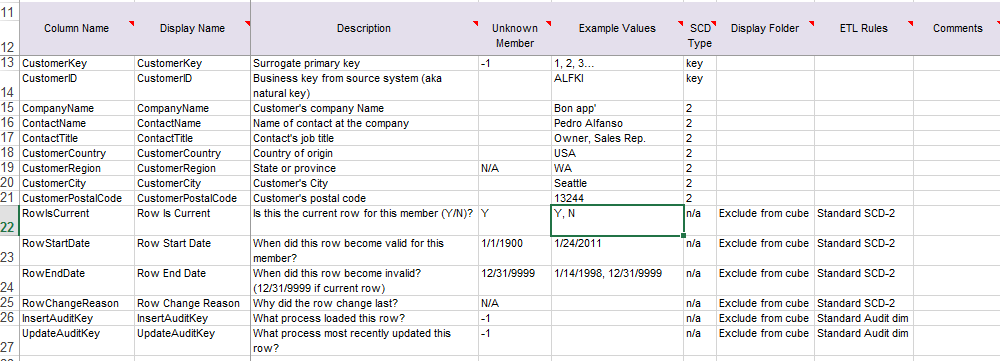
Must match the worksheet tab name for the diagram to work

#### Step 3: Complete the basic column information

Next we complete the basic column definitions using the data from our **Attributes & Measures** from the high level design. Here’s an explanation of the first 9 columns:

* **Column Name** 🡪 Physical name of column in the table
* **Display Name** 🡪 Logical name of column in the table (should match the physical name)
* **Description** 🡪 Explanation of the column, for documentation purposes
* **Unknown Member** 🡪 What should be used for an unknown value (in place of NULL)
* **Example Values** 🡪 What do sample values look like?
* **SCD Type** 🡪 Slowly changing dimension type: key (does not change), 1,2,3, or n/a
* **Display Folder** 🡪 Provides grouping for similar attributes / facts in a cube design.
* **ETL Rules** 🡪 Any special ETL rules, if known at this time.

Fill out your column information to match mine:



**NOTE:** Included in this detailed design are techniques for dealing with type-2 SCD’s and an audit dimension (everything from row 22 and higher in the screenshot). Both of these techniques are covered in the ETL chapters of our course. For now, we can leave these in our design. We’ll revisit them later.

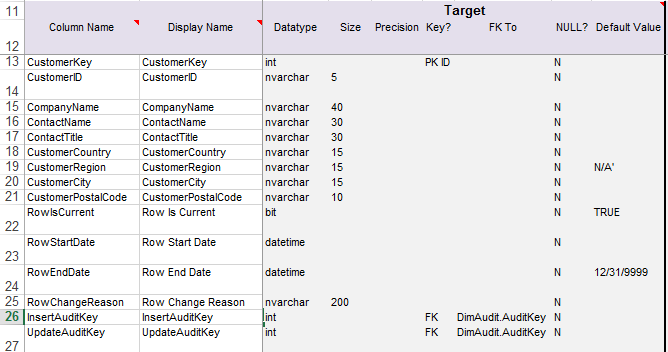
#### Step 4: Target (dimensional table) table definition

In this next step, you’ll define the table definition for our dimension table in our ROLAP star schema. You work here follows along with normal relational table design definitions.

The columns you’ll need to complete in this step for each attribute are:

* **Datatype, Size, Precision** – the SQL Server datatype (including size and precision, where appropriate) of the attribute. A good rule of thumb is to check the source data type for reference. It should be noted that data types vary from DBMS to DBMS. SQL server datatype reference can be found at <http://msdn.microsoft.com/en-us/library/ms187752.aspx>.
* **Key?** – Should be blank if not a key or labeled PK = primary key, PK ID = primary key (with surrogate), or FK = foreign key.
* **FK To** – When you label an attribute as FK, you need to include a dimension table and its primary key as the referencing column. (This is used in the Fact table).
* **NULL**? – Whether or not the attribute permits null values. This should only be permitted in very rare circumstances. The better design decision is to provide a default value in place of NULL.
* **Default Value** – A value which should be stored in the event there is no value.

Here’s a screenshot of my completed target definition for customer:

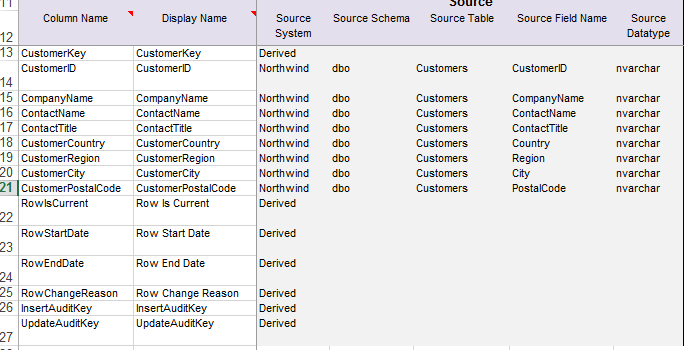


#### Step 5: Source definition

In this final step you’ll complete the source definition, which will assist us as we complete the ETL implementation in a subsequent step. Here’s an explanation of the columns you’ll need to complete in this step:

* **Source System** – List the source system for the attribute. ***Derived*** implies the attribute is calculated.
* **Source Schema** – If the attribute comes from a specific schema, list it here.
* **Source Table** – State the table the attribute comes from on the source system.
* **Source Field Name** – The column or columns which supply the attribute. If the column is a calculation, specify that here (ex. OrderQty\*Price).

Here’s a screenshot of my completed target definition for customer:



# Part 3: On Your Own

Let’s complete the detailed worksheet for the **Sales Reporting** and **Order Fulfillment** business processes. At Minimum, you should have the following tables.

|  |  |
| --- | --- |
| **High Level Bus Matrix** | **Create this in the Detail Level as a Database Table** |
| Customers | DimCustomer (we did this is part 2) |
| Order Date, Ship Date, Etc… | DimDate (done for you already) |
| Employees | DimEmployee |
| Products | DimProduct |
| Sales Reporting | FactSales |
| Order Fulfillment | FactOrderFulfillment |

**NOTE:** You may have others such as Shippers and Suppliers based on your **Bus Matrix** design.

# Turning it in:

Please turn in your completed **Detailed Dimensional Modeling Workbook** with your name, NetID, and date somewhere in the sheet of the **Home** tab.

If you worked with a partner, please indicate that in your assignment by including your partner’s name and NetID. You should both submit the assignment individually.