

# Assignment 04:

## 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, 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 because 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 for your ETL process and contains Excel macros that 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 because there are many details I will not explain because they are covered in the workbook instructions.

### Getting Started

First let's set up the workbook.

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

- Database: **ist722\_yournetid\_dw** (replace *yournetid* with your actual NetID)
- Description: **The Northwind Traders Data Warehouse**
- Gen FKs?: **Y**
- Schema for Views: **(leave blank)**

Here's a screenshot of my completed worksheet.

Date:	Friday, February 18, 2011
Database:	ist722_mafudge_dw
Description:	The Northwind Traders Data Warehouse
Version:	4.00
Gen FKs?:	Y
Schema for Views:	

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 and 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 and Measures** from the high-level design. A screenshot has been included for reference.

	A	B	C	D	E
1	<b>Instructions!</b>				
	Dimension /	Attribute /		Alternate	Sample
2	Fact Table	Fact Name	Description	Names	Values
3	Customer	Company Name	The name of the customer's company		Bon App'
4	Customer	Contact Name	The name of the contact at the company		Thomas Hardy
5	Customer	Contact Title	The contact's title at the company		Owner
6	Customer	Customer Country	Country of origin for the customer		France
7	Customer	Customer Region	State or province for the customer (not aval sometimes)		WA
8	Customer	Customer City	Customer's city		London
9	Customer	Customer Postal Code	Customer's postal code		13244

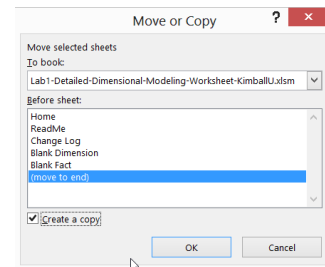
The process you'll follow to design a dimension or fact table is outlined in **five steps**:

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**.

**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:

	A	B	
1	Table Name	DimCustomer	The name of table
2	Table Type	Dimension	
3	Display Name	DimCustomer	Must match the worksheet tab name for the diagram to work
4	Database Schema		
5	Table Description	Customer dimension	
6	Comment	sourced from Northwind	These fields should be completed as necessary
7	Biz Filter Logic		
8	Size	one rows for each customer	
9	Generate Script?	Y	Change to Y

### Step 3: Complete the Basic Column Information

Next we complete the basic column definitions using the data from our **Attributes and Measures** from the high-level design. Here's an explanation of the first nine 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:

	Column Name	Display Name	Description	Unknown Member	Example Values	SCD Type	Display Folder	ETL Rules	Comments
12	CustomerKey	CustomerKey	Surrogate primary key	-1	1, 2, 3...	key			
13	CustomerID	CustomerID	Business key from source system (aka natural key)		ALFKI	key			
14	CompanyName	CompanyName	Customer's company Name		Bon app'	2			
15	ContactName	ContactName	Name of contact at the company		Pedro Alfonso	2			
16	ContactTitle	ContactTitle	Contact's job title		Owner, Sales Rep.	2			
17	CustomerCountry	CustomerCountry	Country of origin		USA	2			
18	CustomerRegion	CustomerRegion	State or province	N/A	WA	2			
19	CustomerCity	CustomerCity	Customer's City		Seattle	2			
20	CustomerPostalCode	CustomerPostalCode	Customer's postal code		13244	2			
21	RowIsCurrent	Row Is Current	Is this the current row for this member (Y/N)?	Y	Y, N	n/a	Exclude from cube	Standard SCD-2	
22	RowStartDate	Row Start Date	When did this row become valid for this member?	1/1/1900	1/24/2011	n/a	Exclude from cube	Standard SCD-2	
23	RowEndDate	Row End Date	When did this row become invalid? (12/31/9999 if current row)	12/31/9999	1/14/1998, 12/31/9999	n/a	Exclude from cube	Standard SCD-2	
24	RowChangeReason	Row Change Reason	Why did the row change last?	N/A		n/a	Exclude from cube	Standard SCD-2	
25	InsertAuditKey	InsertAuditKey	What process loaded this row?	-1		n/a	Exclude from cube	Standard Audit dim	
26	UpdateAuditKey	UpdateAuditKey	What process most recently updated this row?	-1		n/a	Exclude from cube	Standard Audit dim	

**NOTE:** Included in this detailed design are techniques for dealing with type-2 SCDs 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. Your 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 be permitted only in very rare circumstances. The better design decision is to provide a default value in place of NULL.
- **Default Value**—A value that should be stored in the event there is no value.

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

11			Target						
	Column Name	Display Name	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value
12									
13	CustomerKey	CustomerKey	int			PK ID		N	
14	CustomerID	CustomerID	nvarchar	5				N	
15	CompanyName	CompanyName	nvarchar	40				N	
16	ContactName	ContactName	nvarchar	30				N	
17	ContactTitle	ContactTitle	nvarchar	30				N	
18	CustomerCountry	CustomerCountry	nvarchar	15				N	
19	CustomerRegion	CustomerRegion	nvarchar	15				N	N/A
20	CustomerCity	CustomerCity	nvarchar	15				N	
21	CustomerPostalCode	CustomerPostalCode	nvarchar	10				N	
22	RowIsCurrent	Row Is Current	bit					N	TRUE
23	RowStartDate	Row Start Date	datetime					N	
24	RowEndDate	Row End Date	datetime					N	12/31/9999
25	RowChangeReason	Row Change Reason	nvarchar	200				N	
26	InsertAuditKey	InsertAuditKey	int			FK	DimAudit.AuditKey	N	
27	UpdateAuditKey	UpdateAuditKey	int			FK	DimAudit.AuditKey	N	

#### 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 that supply the attribute. If the column is a calculation, specify that here (e.g., OrderQty\*Price).

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

	Column Name	Display Name	Source System	Source Schema	Source Table	Source Field Name	Source Datatype
11							
12							
13	CustomerKey	CustomerKey	Derived				
14	CustomerID	CustomerID	Northwind	dbo	Customers	CustomerID	nvarchar
15	CompanyName	CompanyName	Northwind	dbo	Customers	CompanyName	nvarchar
16	ContactName	ContactName	Northwind	dbo	Customers	ContactName	nvarchar
17	ContactTitle	ContactTitle	Northwind	dbo	Customers	ContactTitle	nvarchar
18	CustomerCountry	CustomerCountry	Northwind	dbo	Customers	Country	nvarchar
19	CustomerRegion	CustomerRegion	Northwind	dbo	Customers	Region	nvarchar
20	CustomerCity	CustomerCity	Northwind	dbo	Customers	City	nvarchar
21	CustomerPostalCode	CustomerPostalCode	Northwind	dbo	Customers	PostalCode	nvarchar
22	RowIsCurrent	Row Is Current	Derived				
23	RowStartDate	Row Start Date	Derived				
24	RowEndDate	Row End Date	Derived				
25	RowChangeReason	Row Change Reason	Derived				
26	InsertAuditKey	InsertAuditKey	Derived				
27	UpdateAuditKey	UpdateAuditKey	Derived				

### Part 3: On Your Own

Let's complete the detailed worksheet for the **Sales Reporting** and **Order Fulfillment** business processes. At a 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 in Part 2)
Order Date, Ship Date, and so on	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.

#### Hints

Be sure to identify the columns that form the primary key with the label PK.

Facts (measures) do not belong in dimension tables.