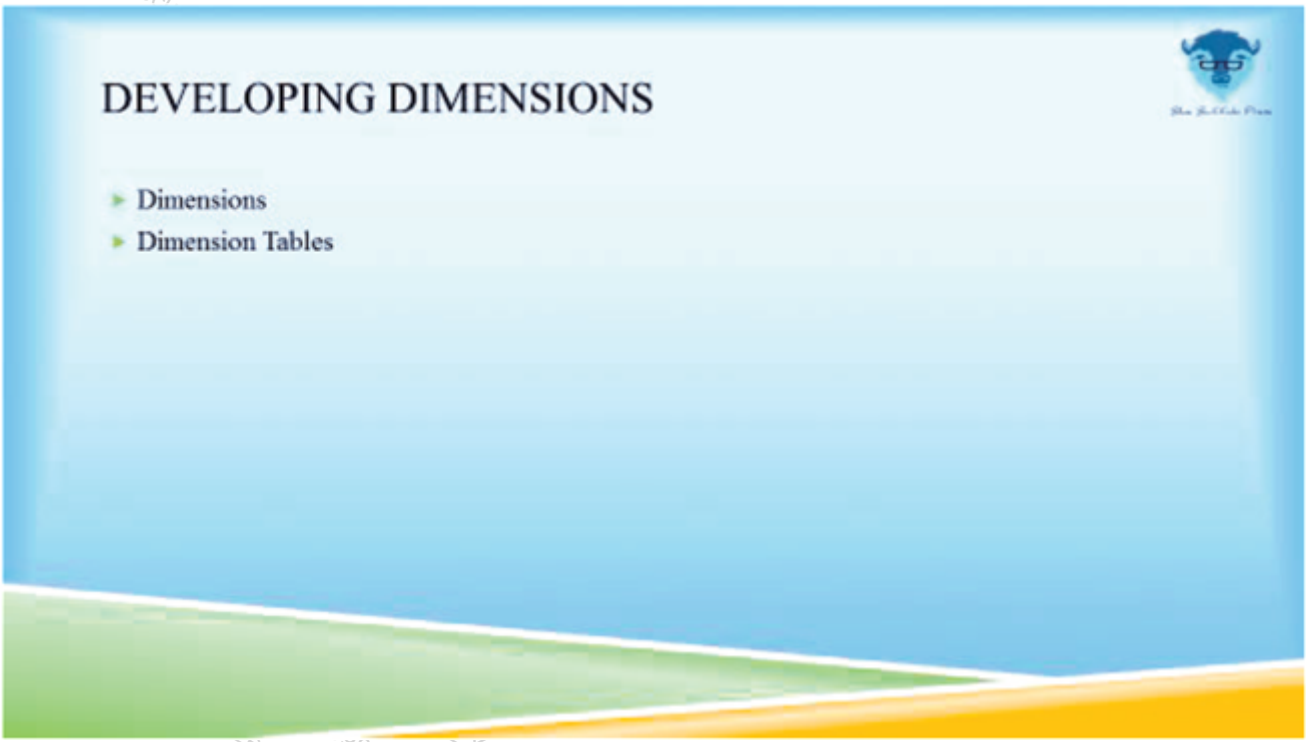
4 DISSECTING WITH DIMENSIONS

MODULE OBJECTIVE

Dimensions are the categories by which you slice your data to view specific quantities of interest. In this module we will explore dimensions and how they are used in conjunction with attributes, relationships, and hierarchies.

MODULE TOPICS

DEVELOPING DIMENSIONS



DIMENSIONS

Dimensions are the "by" that we use to build a frame around our measures.

- Gross Sales BY Product
- Gross Sales BY Industry

► Gross Sales BY Region

You define multidimensional space by specifying a set of dimensions. Therefore, the dimension is the central object of the multidimensional model. Many other objects are based on the dimension.

DIMENSION TABLES

Like Facts, or Measures, Dimensions are stored in dimension tables.

DEMONSTRATION

VIDEO: DEVELOPING DIMENSIONS

EXERCISE A.1: DEVELOPING DIMENSIONS

Objective: in this exercise we will create and configure two dimensions.

A.1.1 Switch back to **SQL Server Management Studio**.

A.1.2 Navigate to the **Object Explorer** pane on the left and move to the **AdventureWorksDW2012** database where **Tables** folder is expanded.

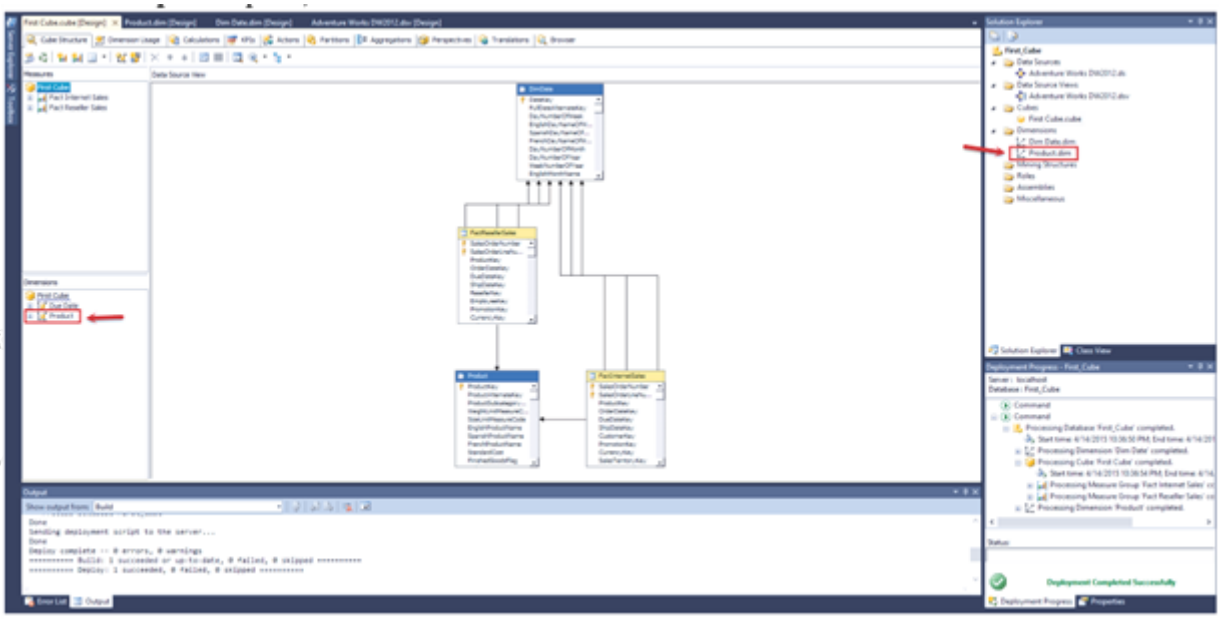
A.1.3 Locate and expand the **dbo.DimProduct** table.

A.1.4 Expand **Columns** folder. Review the results and the listed columns. Expand the pane if necessary, by using drag-and-drop and repositioning the right border.

A.1.5 Switch to **SQL Server Data Tools**.

A.1.6 Back in **Microsoft Visual Studio**, click **Cube Structure** tab.

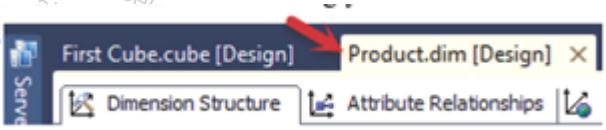
A.1.7 Notice you see **Product** dimension listed on both sides of the interface (**Dimensions** pane, and **Solution Explorer** pane).



A.1.8 Navigate to the **Solution Explorer** pane on the right, move to the **Dimensions** folder, and click to select **Product.dim**.

A.1.9 Right-click **Product.dim**, review the options available, then click **Open**.

A.1.10 Review the results noticing you are now in a new tab (**Product.dim [Design]**).



A.1.11 Move to the **Attributes** section on the left, and click to select **Product** dimension.

A.1.12 Right-click **Product** dimension and review the options available.

A.1.13 Move to the **Data Source View** section and review the table.

A.1.14 Click **First Cube.cube [Design]** tab.

A.1.15 Navigate to the **Dimension** section on the left, right-click **Product** dimension, and click **Edit Dimension**.

A.1.16 Notice you are back at the **Product.dim [Design]** tab.

A.1.17 Move to the **Data Source View** section, locate **Size** within the **Product** table, and click to select it.

A.1.18 Right-click **Size**, review the options available, and click **Explore Data**.

A.1.19 Review the results noticing a new tab opened (**Explore Product Table**). Review the data.

A.1.20 Close the **Explore Product Table** tab.

A.1.21 Back in **Product.dim [Design]** tab, navigate back to the **Data Source View** section, locate **Size** within the **Product** table, and click to select it.

A.1.22 Right-click **Size** again, and click **New Attribute from Column**.

A.1.23 Review the results in the **Attributes** section noticing **Size** is now listed.

A.1.24 Move to the **Attributes** section, click to select **Size** attribute, then right-click **Size**, review the options available, and then click **Properties**.

A.1.25 Notice the **Properties** pane appeared in the lower-right. Review the properties.

A.1.26 Navigate up to the toolbar and click  (**Save All**).

A.1.27 Move to the **Solution Explorer** pane on the right, right-click the **First_Cube** project, then click **Deploy**.

A.1.28 Notice in the lower-left you see **Deploy succeeded**.

A.1.29 Click **First Cube.cube [Design]** tab.

A.1.30 Click **Browser** tab.

A.1.31 When the browser opens, move to the **Metadata** pane on the left, and scroll down to the dimensions.

A.1.32 Expand **Product** dimension, and notice **Size** is listed.

If you don't see Size listed, click  *Reconnect.*

A.1.33 Scroll up to the **Measures**, and expand if necessary.

A.1.34 Expand **Fact Internet Sales** folder.

A.1.35 Locate **Sales Amount**, right-click the measure, and click **Add to Query**.

Review the results.

A.1.36 Scroll down to **Product** dimension.

A.1.37 Locate **Size**, right-click the attribute, and then click **Add to Query**. Review the results.

A.1.38 Move back to the **Metadata** pane on the left, and expand **Size** attribute. Review the results

A.1.39 Expand **Members** folder. Review the results.

A.1.40 Expand **All**. Review the results.

A.1.41 Navigate up to the toolbar and click  (**Save All**).

A.1.42 Close all open windows.

UTILIZING ATTRIBUTES

DIMENSION ATTRIBUTES

The multidimensional model contains a collection of attributes that defines a set of "domains" of a dimension's data; one domain is a dimension attribute. We use the

term domain as it is defined in the relational database theory: a limited list of values (scalars). However, implementations of relational databases essentially ignore this definition. In practice, implementations of the relational data model manipulate columns that will accept any value (of the appropriate type).

In the multidimensional model, attribute defines domain—a list of the same type values, which we call key values. All key values of attribute have the same data type. Key is a unique identifier of a dimension member.

An attribute is the full collection of members. For example, all the days of the week would be an attribute of the Time dimension.

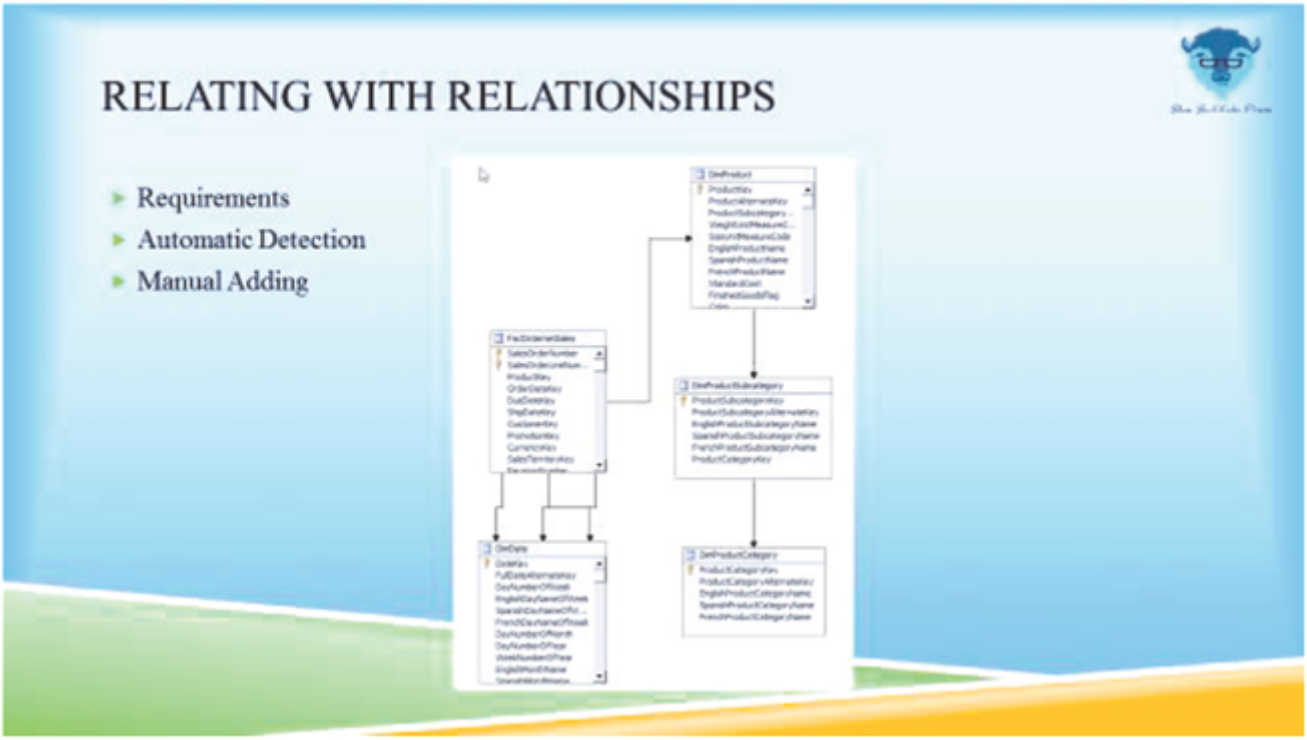
Cardinality of Dimension

The size, or cardinality, of a dimension is the number of members it contains. For example, a Time dimension made up of the days of the week would have a size of 7.

MEMBERS

A member corresponds to one point on a dimension. For example, in the Time dimension, Monday would be a dimension member. A value is a unique characteristic of a member. For example, in the Time dimension, 5/12/2008 might be the value of the member with the caption "Monday".

RELATING WITH RELATIONSHIPS



REQUIREMENTS

Multiple relationships could result in ambiguous dependencies between tables. To create accurate calculations, you need a single path from one table to the next. Therefore, there can be only one active relationship between each pair of tables.

A source column cannot participate in multiple relationships. If you have used a column as a source column in one relationship already, but want to use that column to connect to another related lookup column in a different table, you can create a copy of the column, and use that column for the new relationship.

The data types in the source column and lookup column must be compatible.

The data values in the lookup column must be unique. In other words, the column cannot contain duplicates. In tabular models, nulls and empty strings are equivalent to a blank, which is a distinct data value. This means that you cannot have multiple nulls in the lookup column.

Each table must have a single column that uniquely identifies each row in that table. This column is often referred to as the primary key.

You cannot use composite keys in a tabular model; you must always have one column that uniquely identifies each row in the table. If you try to import tables that

have an existing relationship based on a composite key, the Table Import Wizard will ignore that relationship because it cannot be created in the tabular model.

Self-joins are not permitted in tabular model tables. A self-join is a recursive relationship between a table and itself. Self-joins are often used to define parent-child hierarchies.

Tabular models do not support many-to-many relationships, and you cannot add junction tables in the model designer.

AUTOMATIC DETECTION

When you import from a relational data source table, the Table Import Wizard detects existing relationships in those source tables based on the source schema data. If related tables are imported, those relationships will be duplicated in the model.

MANUAL ADDING

If your model contains data from multiple sources, you will likely have to manually create relationships.

DEMONSTRATION

VIDEO: RELATING WITH RELATIONSHIPS

EXERCISE B.1: RELATING WITH RELATIONSHIPS

Objective: in this exercise we will connect to data and review the existing relationships, then we'll add a table using the relationships.

B.1.1 Navigate down to the taskbar, right-click **SQL Server Data Tools** icon, right-click the new **SQL Server Data Tools** icon showing, and then click **Run as administrator**.

B.1.2 In the **User Account Control** dialog box, click **Yes**.

B.1.3 When **Microsoft Visual Studio** opens, navigate to the **Recent Project** section, and click to open **First_Cube**.

B.1.4 When **First_Cube** opens, move up to the menu, and click **File | New | Project…**.

B.1.5 When the **New Project** dialog box opens, navigate to the left side of the dialog box, in the **Installed Templates** section, verify **Business Intelligence** is selected, then move to the middle of the dialog box, and click to select **Analysis Services Multidimensional and Data Mining Project**.

B.1.6 Navigate down to the **Solution** setting, use the corresponding drop-down arrow, and click to select **Add to solution**.

B.1.7 Move to the **Name** text box and change the name to Second Cube Related.

B.1.8 Click **OK**.

B.1.9 Navigate to **Solution Explorer** pane on the right, and review the results.

B.1.10 Collapse **First_Cube** project so you can better see **Second Cube Related** project.

B.1.11 Right-click **Data Sources** folder and click **New Data Source…**.

B.1.12 In the **Welcome to the Data Source Wizard** dialog box, review the welcome message, then click **Next**.

B.1.13 When the **Select how to define the connection** dialog box opens, review the options and settings available.

B.1.14 Notice the current connection into **AdventureWorksDW2012** is selected, and then click **Finish**.

B.1.15 In the **Completing the Wizard** dialog box, review the settings.

B.1.16 Click **Finish**.

B.1.17 Navigate back to **Solution Explorer** pane on the right and review the results in the **Data Sources** folder.

B.1.18 Right-click **Data Source Views** folder, and click **New Data Source View…**.

B.1.19 In the **Welcome to the Data Source View Wizard** dialog box, review the welcome message, then click **Next**.

B.1.20 In the **Select a Data Source** dialog box, review the settings.

B.1.21 Click **Next**.

B.1.22 When the **Select Tables and Views** dialog box opens, review the settings and options available.

B.1.23  Maximize the window.

B.1.24 Navigate to the **Available objects** section and click to select **DimProduct (dbo)**.

B.1.25 Click  and notice **DimProduct (dbo)** is now listed in the **Included objects** section.

B.1.26 Click **Add Related Tables**.

B.1.27 Review the results noticing **FactInternetSales (dbo)**, **FactResellerSales (dbo)**, **FactProductInventory (dbo)**, and **DimProductSubcategory (dbo)** were added.

B.1.28 Move to the **Included objects** section, click to select **FactResellerSales (dbo)**, and then click .

B.1.29 Notice FactResellerSales (dbo) is no longer listed in the **Included objects** section.

B.1.30 Click to select **FactProductInventory (dbo)**, and then click .

B.1.31 Notice FactProductInventory (dbo) is no longer listed in the **Included objects** section.

B.1.32 Click to select **DimProductSubcategory (dbo)**, and then click **Add Related Tables**.

B.1.33 Review the results noticing **DimProductCategory (dbo)** was added.

B.1.34 Navigate to the **Available objects** section and click to select **DimDate (dbo)**.

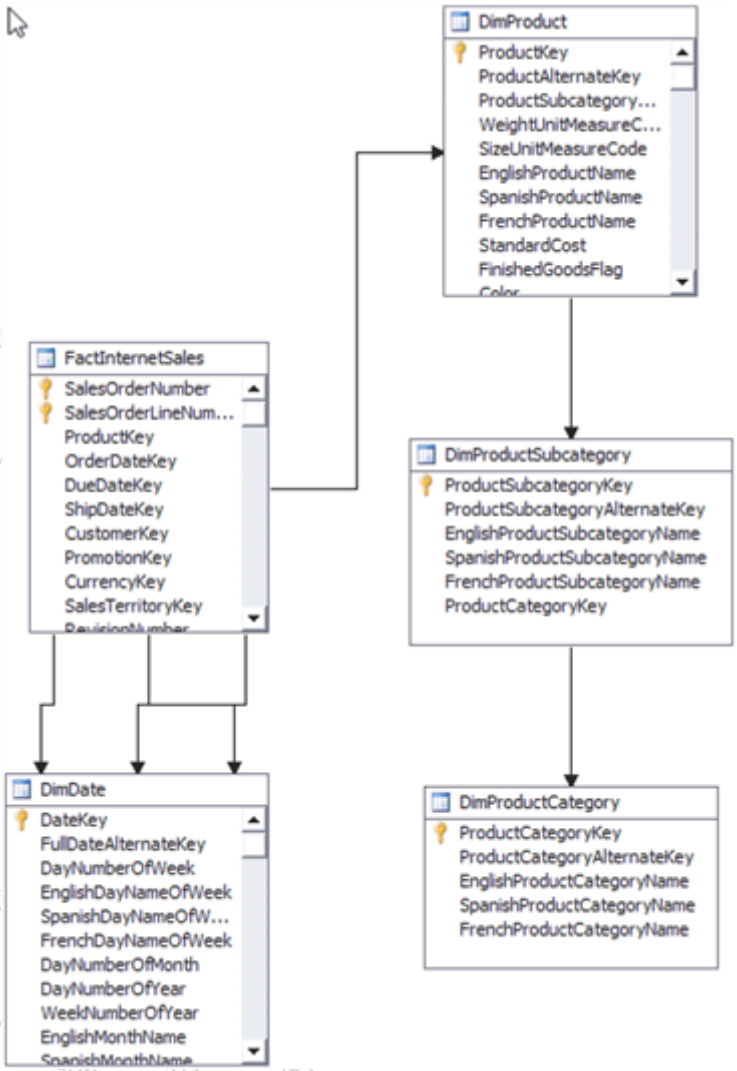
B.1.35 Click  and notice **DimDate (dbo)** is now listed in the **Included objects** section.

B.1.36 Click **Next**.

B.1.37 In the **Completing the Wizard** dialog box, review the settings.

B.1.38 Click **Finish**.

B.1.39 Review the results noticing the tables and their existing connections.



B.1.40 Move to the design area, click to select **DimProductCategory** table, right-click that same table, and click **Delete Table From DSV**.

B.1.41 In the **Delete Objects** dialog box advising **The following objects will be deleted**, click **OK**.

B.1.42 Review the results noticing the DimProductCategory table is no longer listed.

B.1.43 Right-click a blank spot within the design area and then click **Add/Remove Tables…**.

B.1.44 When the **Add/Remove Tables** dialog box opens, review the options available.

B.1.45 Move to the **Included objects** section, click to select **DimProductSubcategory (dbo)**, and then click **Add Related Tables**.

B.1.46 Review the results noticing **DimProductCategory (dbo)** was added to the **Included objects** section.

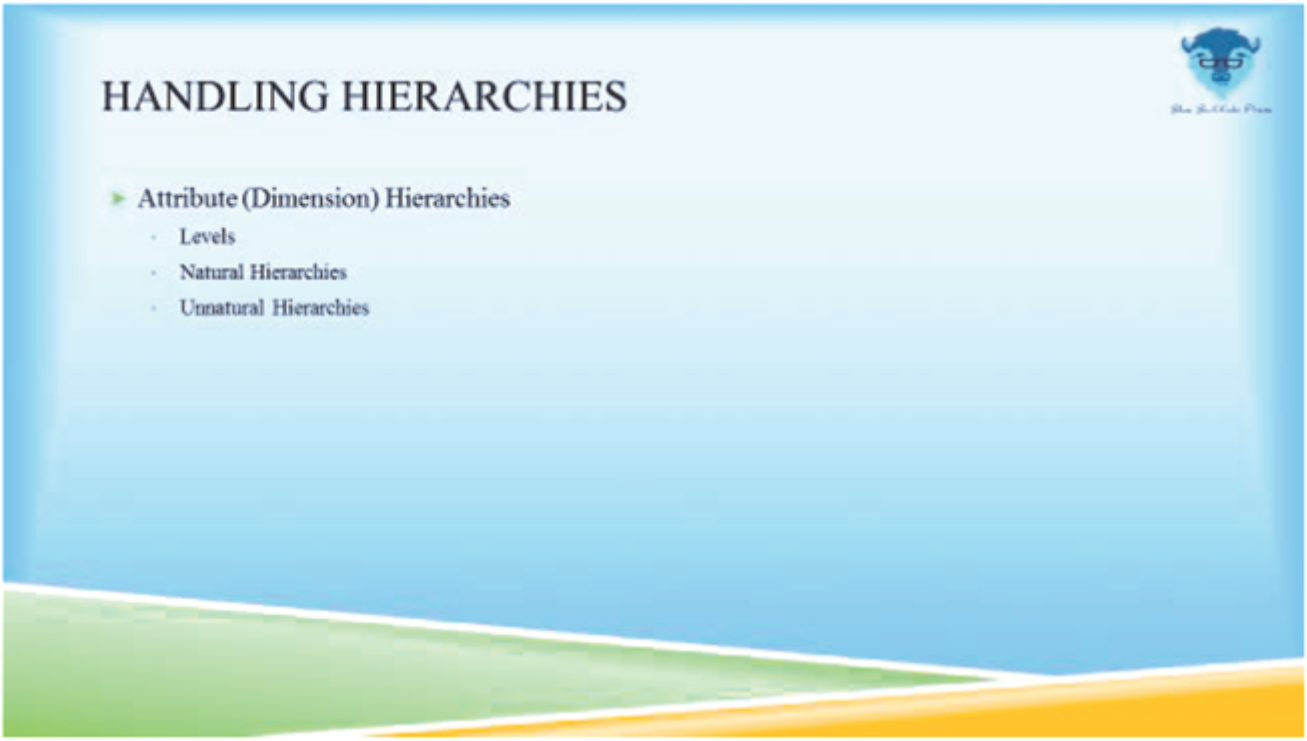
B.1.47 Click **OK**.

B.1.48 Review the results noticing you now see **DimProductCategory** in the design area.

B.1.49 Navigate up to the toolbar and click  (**Save All**).

B.1.50 Close all open windows.

HANDLING HIERARCHIES



ATTRIBUTE (DIMENSION) HIERARCHIES

The hierarchy, which can be seen as a navigational path, is the primary means of access to the multidimensional model. It consists of attributes arranged in descending (or ascending, depending on the way you look at it) levels, with each level containing attribute members. Efficient use of hierarchies significantly increases the effectiveness of the multidimensional model.

Levels

The task of defining a hierarchy is simply to define the hierarchy object itself, and to specify the collection of all the levels and all their properties. Levels have two

important properties:

- SourceAttributeID— defines the source for the level members. After the SourceAttributeID has defined a member as belonging to a level, the member gains two characteristics:
  - A parent member is a member from the level above, which current member belongs to.
  - Children are members from the next level that belongs to the current member.

  This gives us the ability to navigate the hierarchy.
- HideMemberIf— defines the rules for displaying members for the client application. Some members will be hidden so that not all the members of the source attribute will be apparent in the hierarchy. Use of this property significantly diminishes the effectiveness of the hierarchy and complicates the calculation algorithms.

Natural Hierarchies

- In the Customer dimension, dependencies exist between attributes. Because country is a dependent attribute of the state, if you know the state, you can unambiguously determine what country it belongs to. The same is true for city and state. Because the state is a related attribute for the city, if you know the city you can unambiguously say what state it belongs to.
- In this hierarchy, the presence of a key attribute changes nothing. The entire structure of the hierarchy is determined by the relationships between the attributes that are used as levels in the hierarchy. Such hierarchy is called a natural hierarchy. All levels of a natural hierarchy are built from related attributes, and the levels are located in correspondence with the direction of the relationships of attributes. As its name suggests, it is the most effective form of hierarchy. In Analysis Services 2000, this was the only possible type of hierarchy.

Unnatural Hierarchies

Analysis Services 2005 introduced a new type of hierarchy, the unnatural hierarchy. As an example, Marital Status and Gender, which are not related to each other. If you used an earlier version of Analysis Services to create a two-level hierarchy that includes these two attributes, you would have to perform a whole series of crafty manipulations with the data. With the current version of Analysis Services, it is just as easy to create this type of hierarchy as it was in earlier versions to create a natural hierarchy.

- The unnatural hierarchy differs from the natural hierarchy in the way it defines which members belong to which parents. Whereas the natural hierarchy defines its parent/child relationships by the relationships between members, the unnatural hierarchy defines its parent/child relationships by the relationships of the members to the key attribute.

- Another difference between the two types of hierarchies is that the unnatural hierarchy can support many-to-many relationships among the members from different levels of the hierarchy. The natural hierarchy, in contrast, can support only a one-to-many relationship. So, in the Customer dimension, with an unnatural hierarchy the attribute member with Marital Status "married" could have two parents, Male and Female, as could the attribute member "unmarried".

- At first glance, natural and unnatural hierarchies appear to have the same look and same behavior. However, in multidimensional space, as in life, looks are deceiving. It's best to use natural hierarchy unless you really need an unnatural hierarchy. Because of the types of relationships between attributes, you could see a drop in performance and slow performance may kill a project!

DEMONSTRATION

VIDEO: HANDLING HIERARCHIES

EXERCISE C.1: HANDLING HIERARCHIES

Objective: in this exercise we will create and configure a hierarchy.

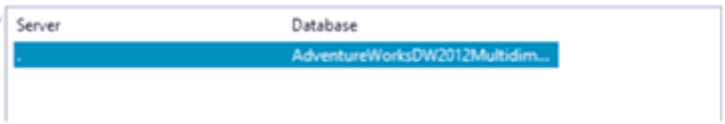
C.1.1 Navigate down to the taskbar, right-click **SQL Server Data Tools** icon, right-click the new **SQL Server Data Tools** icon showing, and then click **Run as administrator**.

C.1.2 In the **User Account Control** dialog box, click **Yes**.

C.1.3 When **Microsoft Visual Studio** opens, navigate to the upper-left, click **File | Open | Analysis Services Database…**.

C.1.4 In the **Connect To Database** dialog box, review the settings and options available.

C.1.5 Navigate to the box below the **Database** setting, and click to select the existing connection into **AdventureWorksDW2012Multidimensional-EE**.



C.1.6 Click **OK**.

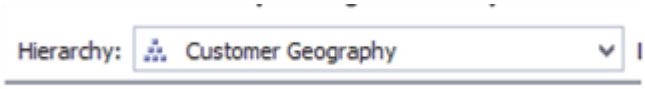
C.1.7 Move to **Solution Explorer** pane on the right, and review the results.

C.1.8 Navigate to the **Dimensions** folder, and double-click **Customer** dimension. Review the results.

C.1.9 Move to the **Hierarchies** section in the center, and review the existing hierarchy.

C.1.10 Click **Browser** tab. Review the results.

C.1.11 Notice the **Hierarchy** setting is currently set to **Customer Geography**.



C.1.12 Expand **All Customers**. Review the results.

C.1.13 Expand **United States**. Review the results.

C.1.14 Expand **Michigan**. Review the results.

C.1.15 Expand **Detroit**. Review the results.

C.1.16 Expand **Howell**. Review the results.

C.1.17 Click **Dimension Structure** tab.

C.1.18 Move back to the **Hierarchies** section in the center, and review the hierarchy structure noting each level.

C.1.19 Click **Attribute Relationships** tab. Review the results.

C.1.20 Click **Dimension Structure** tab.

C.1.21 Navigate to the **Attributes** section on the left, right-click **City** and click **Start New Hierarchy**.

C.1.22 Review the results in the **Hierarchies** section in the center.

C.1.23 Move back to the **Attributes** section on the left, and using drag-and-drop, take **Customer** and drop it into the (newly created) **Hierarchy** box in the center, onto the **<new level>** cell. Review the results.

C.1.24 Right-click the header cell of the **Hierarchy** box in the center, and click **Rename**.

C.1.25 Enter Customer by City.

C.1.26 Click a blank spot in the design area and review the results.

C.1.27 Right-click the header cell of the **Customer by City** box in the center, and click **Delete**. Review the results.

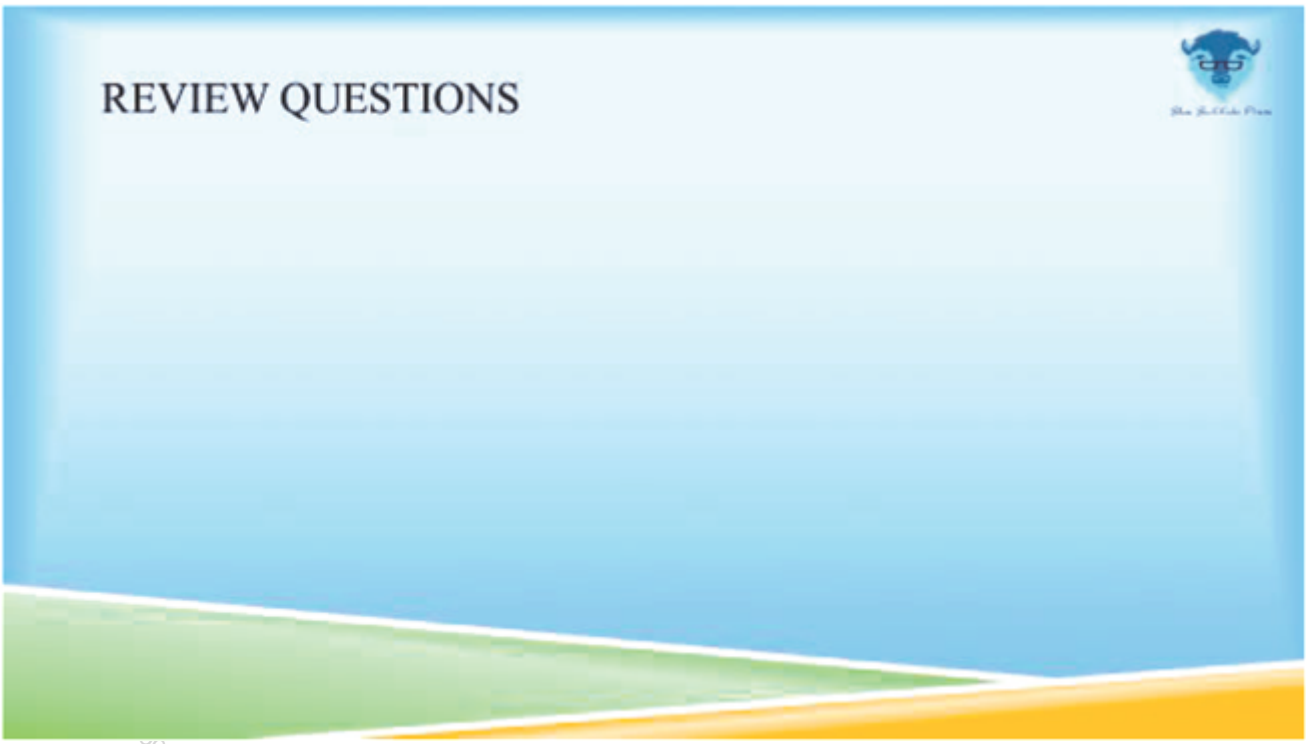
C.1.28 Close all open windows and when prompted to save changes, click **No**.

MODULE REVIEW

MODULE OBJECTIVE

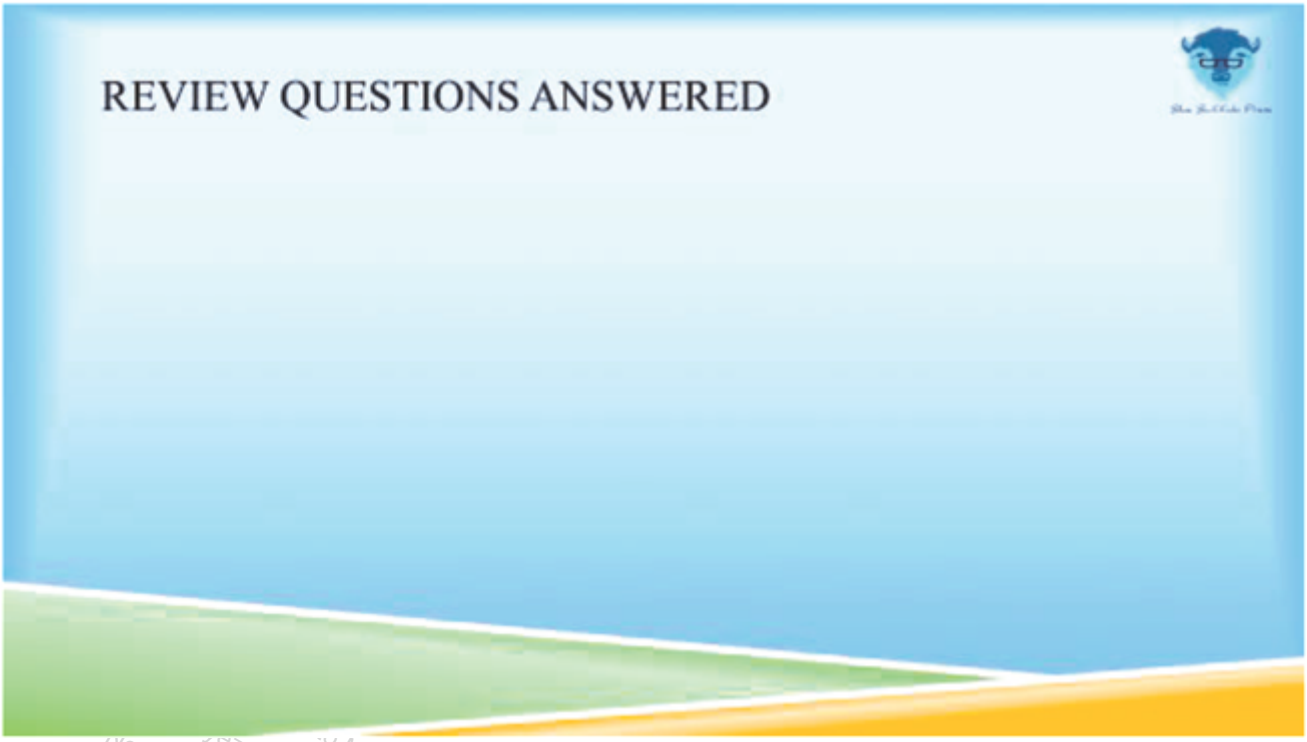
Dimensions are the categories by which you slice your data to view specific quantities of interest. In this module we will explore dimensions and how they are used in conjunction with attributes, relationships, and hierarchies.

REVIEW QUESTIONS



1. You define multidimensional space by specifying a set of _____.
2. The multidimensional model contains a collection of _____ that defines a set of "domains" of a dimension's data.
3. The size, or_____, of a dimension is the number of members it contains.
4. A _____ corresponds to one point on a dimension.
5. The _____, which can be seen as a navigational path, is the primary means of access to the multidimensional model.

REVIEW QUESTIONS ANSWERED



1. You define multidimensional space by specifying a set of _____.
   a. Dimensions
2. The multidimensional model contains a collection of _____ that defines a set of "domains" of a dimension's data.
   a. Attributes

3. The size, or_____, of a dimension is the number of members it contains.

    a. Cardinality

4. A _____ corresponds to one point on a dimension.

    a. Member

5. The _____, which can be seen as a navigational path, is the primary means of access to the multidimensional model.

    a. Hierarchy