

# Carlo Carandang - Assignment 1, DBAS2103 - Creating Source-to-Target Mapping Documents

Printed On: 2017-05-25 10:44

**Due Date: February 9, 2017 5:00 pm**

| Version | Description   | Changed By      | Date       |
|---------|---------------|-----------------|------------|
| 0.1     | First draft   | Carlo Carandang | 2017-02-07 |
| 1       | Revised draft | Carlo Carandang | 2017-02-08 |
|         |               |                 |            |

| Reviewed by    | Date       |
|----------------|------------|
| Michael Nisbet | 2017-02-08 |
|                |            |

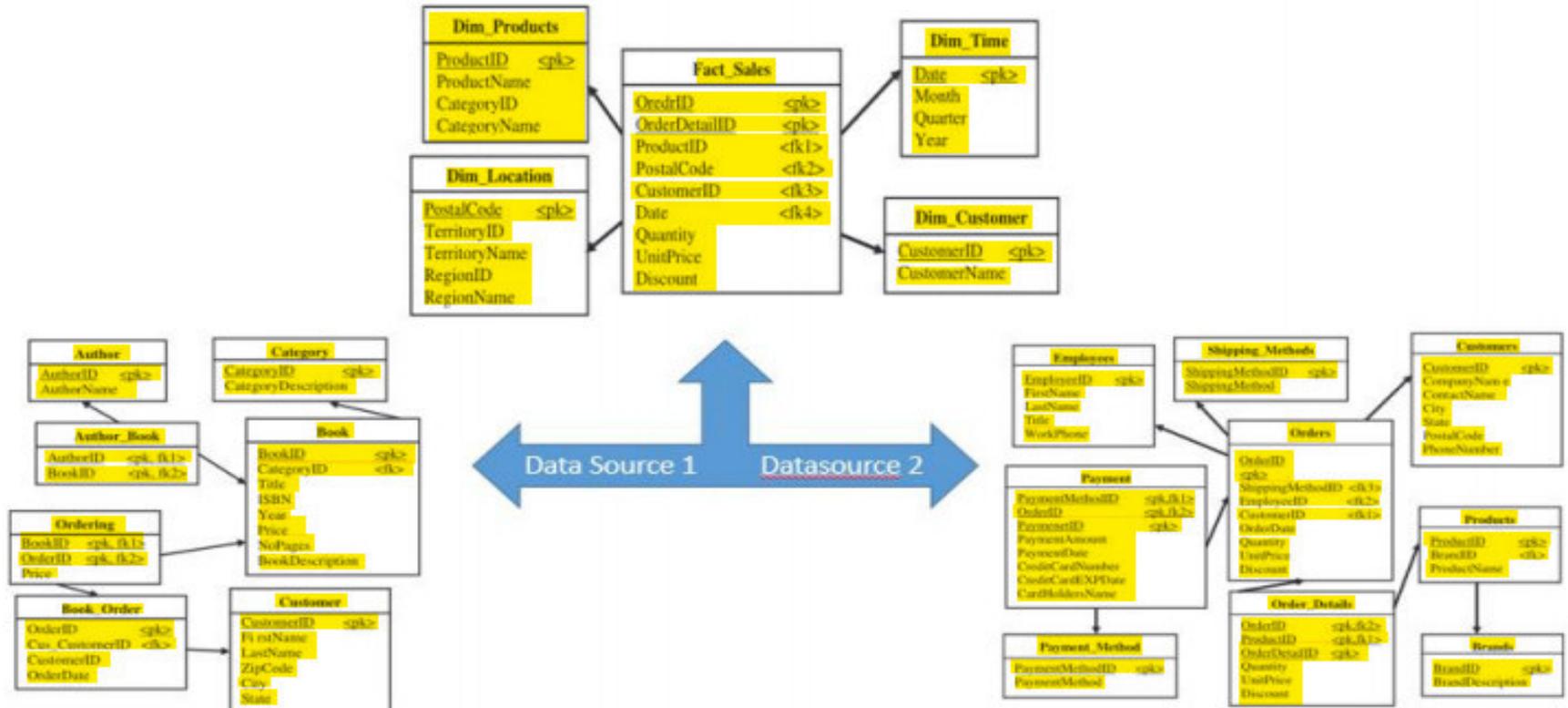
Contents: 11 sheets

## Table of Contents

|                         |          |
|-------------------------|----------|
| Project Purpose.....    | Sheet 3  |
| Dim_Products Table..... | Sheet 4  |
| Dim_Location Table..... | Sheet 5  |
| Dim_Time Table.....     | Sheet 6  |
| Dim_Customer Table..... | Sheet 7  |
| Fact_Sales Table.....   | Sheet 8  |
| Dim_Location Table..... | Sheet 9  |
| Orphan Source Data..... | Sheet 10 |
| Peer Review.....        | Sheet 11 |

1. Purpose. This project entails the creation of a Source-to-Target Mapping Document using MS Excel
2. Project. This project provides detailed ETL specifications for data extraction, transformation, and reconciliation
3. Details. Will map the process of extracting data from 2 sources (Data Source 1 and 2), and transform and load it into the Proposed Star Schema in a Data Warehouse, as outlined below

# Proposed Star Schema in a Data Warehouse



Source: <http://www.christoolivier.com/why-source-to-target-mapping-documents-matter/>

|                            |   |
|----------------------------|---|
| <b>Project</b>             | <i>Proposed Star Schema in a Data Warehouse</i>   |
| <b>ETL Package Name</b>    | <i>Load Dim_Products.dtsx</i>   |
| <b>Target Table</b>        | <i>Dim_Products</i>   |
| <b>Source Tables</b>       | <i>Products<br/>Category</i>  |
| <b>Source Data</b>         | <i>Database to be loaded into Oracle</i>  |
| <b>Target DBMS</b>         | <i>Oracle</i>   |
| <b>Mapping Description</b> |   |
| <b>Error Handling</b>      | On an error occuring any inserted records or updates needs to roll back and leave the system in the |

**Mapping Logic:**

| Source System/ Table  | Source Column       | Source Type   | Target Table | Target Column | Target Datatype | Format | Example       | Constraint | Transformation/Business rule                  | Comments / Other rules         |
|-----------------------|---------------------|---|--------------|---------------|-----------------|--------|---------------|------------|---|--------------------------------|
| Datasource2/ Products | ProductID           | Is this a table, schema, node, column, row, record? | Dim_Products | ProductID     | Number          | #####  | 12345         | PK         | Sequence numbers are generated automatically. | Is this a varchar or a number? |
| Datasource2/ Products | ProductName         | Is this a table, schema, node, column, row, record? | Dim_Products | ProductName   | Varchar(50)     | ABC123 | Book Title 1  |            |   |                                |
| Datasource1/ Category | CategoryID          | Is this a table, schema, node, column, row, record? | Dim_Products | CategoryID    | Number          | #####  | 12345         |            |   | Is this a varchar or a number? |
| Datasource1/ Category | CategoryDescription | Is this a table, schema, node, column, row, record? | Dim_Products | CategoryName  | Varchar(50)     | ABC123 | Category Name |            |   |                                |

Source: <http://www.christoolivier.com/why-source-to-target-mapping-documents-matter/>

|                            |   |
|----------------------------|---|
| <b>Project</b>             | <i>Proposed Star Schema in a Data Warehouse</i>   |
| <b>ETL Package Name</b>    | <i>Load_Dim_Location.dtsx</i>   |
| <b>Target Table</b>        | <i>Dim_Location</i>   |
| <b>Source Tables</b>       | <i>Customers</i>  |
| <b>Source Data</b>         | <i>Database to be loaded into Oracle</i>  |
| <b>Target DBMS</b>         | <i>Oracle</i>   |
| <b>Mapping Description</b> |   |
| <b>Error Handling</b>      | On an error occuring any inserted records or updates needs to roll back and leave the system in the |

**Mapping Logic:**

| Source System/ Table   | Source Column | Source Type   | Target Table | Target Column | Target Datatype | Format  | Example        | Constraint | Transformation/Business rule                  | Comments / Other rules  |
|------------------------|---------------|---|--------------|---------------|-----------------|---------|----------------|------------|---|---|
| Datasource2/ Customers | PostalCode    | Is this a table, schema, node, column, row, record? | Dim_Location | PostalCode    | Varchar(7)      | A#A #A# | B3K 2R4        | PK         |   | Postal Code. Clarifying if this is for Canadian postal codes only, or for postal codes and zip codes for any country? For this exercise, the Canadian postal code will be used as the primary key for the target table. |
| see comments           | see comments  |   | Dim_Location | TerritoryID   | Number          | ###     | 123            |            | Sequence numbers are automativally generated. | For this table, TerritoryID is determined by the postal code. Is this a number or varchar?  |
| see comments           | see comments  |   | Dim_Location | TerritoryName | Varchar(25)     | ABCD    | Territory Name |            |   | For this table, TerritoryName is determined by postal code.   |
| see comments           | see comments  |   | Dim_Location | RegionID      | Number          | ###     | 123            |            | Sequence numbers are automativally generated. | For this table, RegionID is determined by the postal code. Is this a number or varchar?   |
| see comments           | see comments  |   | Dim_Location | RegionName    | Varchar(25)     | ABCD    | Region Name    |            |   | For this table, RegionName is determined by the postal code. Question: is region name the same as territory name?   |

Source: <http://www.christoolivier.com/why-source-to-target-mapping-documents-matter/>

|                            |   |
|----------------------------|---|
| <b>Project</b>             | <i>Proposed Star Schema in a Data Warehouse</i>   |
| <b>ETL Package Name</b>    | <i>Load Dim_Time.dtsx</i>   |
| <b>Target Table</b>        | <i>Dim_Time</i>   |
| <b>Source Tables</b>       | <i>Orders</i>   |
| <b>Source Data</b>         | <i>Database to be loaded into Oracle</i>  |
| <b>Target DBMS</b>         | <i>Oracle</i>   |
| <b>Mapping Description</b> |   |
| <b>Error Handling</b>      | On an error occuring any inserted records or updates needs to roll back and leave the system in the |

**Mapping Logic:**

| Source System/ Table    | Source Column | Source Type   | Target Table | Target Column | Target Datatype | Format        | Example  | Constraint | Transformation/Business rule | Comments / Other rules   |
|-------------------------|---------------|---|--------------|---------------|-----------------|---------------|----------|------------|------------------------------|--|
| Datasource2/ Orders     | OrderDate     | Is this a table, schema, node, column, row, record? | Dim_Time     | Date          | Date            | MMDDYYYY<br>Y | 12012017 | PK         |                              | Which of the to data sources are we using for date? Also need to identify format.  |
| Datasource1/ Book_Order | OrderDate     | Is this a table, schema, node, column, row, record? | Dim_Time     | Date          | Date            | MMDDYYYY<br>Y | 12012017 | PK         |                              | Which of the to data sources are we using for date? Also need to identify format.  |
| see comments            | see comments  |   | Dim_Time     | Month         | Number          |               |          |            |                              | Month determined from Date colum in Dim_Time table, in format MM   |
| see comments            | see comments  |   | Dim_Time     | Quarter       | Varchar(20)     |               |          |            |                              | Quarter determined from Month column in Dim_Time table, where MM 01 to 03 = 1st Quarter, 04 to 06 = 2nd Quarter, 07 to 9 = 3rd Quarter, and 10 to 12 = 4th Quarter |
| see comments            | see comments  |   | Dim_Time     | Year          | Number          |               |          |            |                              | Year determined from Date colum in Dim_Time table, in format YYYY  |

Source: <http://www.christoolivier.com/why-source-to-target-mapping-documents-matter/>

|                            |  |
|----------------------------|--|
| <b>Project</b>             | <i>Proposed Star Schema in a Data Warehouse</i>  |
| <b>ETL Package Name</b>    | <i>Load Dim Customer.dtsx</i>  |
| <b>Target Table</b>        | <i>Dim_Customer</i>  |
| <b>Source Tables</b>       | <i>Customers<br/>Customer</i>  |
| <b>Source Data</b>         | <i>Database to be loaded into Oracle</i>   |
| <b>Target DBMS</b>         | <i>Oracle</i>  |
| <b>Mapping Description</b> | <p>Join the source tables in the following manner:</p> <pre> FROM Datasource1/ Customer INNER JOIN FirstName ON Lastname = CustomerName                     </pre> |
| <b>Error Handling</b>      | On an error occuring any inserted records or updates needs to roll back and leave the system in the  |

**Mapping Logic:**

| Source System/ Table   | Source Column | Source Type   | Target Table | Target Column | Target Datatype | Format              | Example          | Constraint | Transformation/Business rule                  | Comments / Other rules   |
|------------------------|---------------|---|--------------|---------------|-----------------|---------------------|------------------|------------|---|--|
| Datasource2/ Customers | CustomerID    | Is this a table, schema, node, column, row, record? | Dim_Customer | CustomerID    | Number          | ####                | 1234             | PK         | Sequence numbers are generated automatically. | Is this a number or varchar? Which of these two data sources are we using for CustomerID?  |
| Datasource1/ Customer  | CustomerID    | Is this a table, schema, node, column, row, record? | Dim_Customer | CustomerID    | Number          | ####                | 1234             | PK         | Sequence numbers are generated automatically. | Is this a number or varchar? Which of these two data sources are we using for CustomerID?  |
| Datasource2/ Customers | ContactName   | Is this a table, schema, node, column, row, record? | Dim_Customer | CustomerName  | Varchar(50)     | LastName, FirstName | Carandang, Carlo |            |   | Need to identify format, if this is upper or lower case, upper case for initial letter. Which of these two data sources are we using for CustomerName? |
| Datasource1/ Customer  | CustomerName  | Is this a table, schema, node, column, row, record? | Dim_Customer | CustomerName  | Varchar(50)     | LastName, FirstName | Carandang, Carlo |            |   | Need to identify format, if this is upper or lower case, upper case for initial letter. Which of these two data sources are we using for CustomerName? |

Source: <http://www.christoolivier.com/why-source-to-target-mapping-documents-matter/>

|                            |  |
|----------------------------|--|
| <b>Project</b>             | <i>Proposed Star Schema in a Data Warehouse</i>  |
| <b>ETL Package Name</b>    | <i>Load_Fact_Sales.dtsx</i>  |
| <b>Target Table</b>        | <i>Fact_Sales</i>  |
| <b>Source Tables</b>       | <i>Book_Order<br/>Orders</i>   |
| <b>Source Data</b>         | <i>Database to be loaded into Oracle</i>   |
| <b>Target DBMS</b>         | <i>Oracle</i>  |
| <b>Mapping Description</b> |  |
| <b>Error Handling</b>      | On an error occurring any inserted records or updates needs to roll back and leave the system in |

**Mapping Logic:**

| Source System/ Table       | Source Column | Source Type   | Target Table | Target Column | Target Datatype | Format   | Example  | Constraint                         | Transformation/Business rule                  | Comments / Other rules  |
|----------------------------|---------------|---|--------------|---------------|-----------------|----------|----------|------------------------------------|---|---|
| Datasource1/ Book_Order    | OrderID       | Is this a table, schema, node, column, row, record? | Fact_Sales   | OrderID       | Number          | ####     | 1234     | PK                                 | Sequence numbers are generated automatically. | Is this a number or varchar? Which of these two data sources are we using for OrderID?  |
| Datasource2/ Orders        | OrderID       | Is this a table, schema, node, column, row, record? | Fact_Sales   | OrderID       | Number          | ####     | 1234     | PK                                 | Sequence numbers are generated automatically. | Is this a number or varchar? Which of these two data sources are we using for OrderID?  |
| Datasource2/ Order_Details | OrderDetailID | Is this a table, schema, node, column, row, record? | Fact_Sales   | OrderDetailID | Number          | ####     | 1234     | PK                                 | Sequence numbers are generated automatically. | Is this a number or varchar?  |
| Datasource2/ Order_Details | Quantity      | Is this a table, schema, node, column, row, record? | Fact_Sales   | Quantity      | Number          | ####     | 1234     |                                    |   |   |
| Datasource2/ Order_Details | UnitPrice     | Is this a table, schema, node, column, row, record? | Fact_Sales   | UnitPrice     | Number          | \$       | \$400.49 |                                    |   |   |
| Datasource2/ Order_Details | Discount      | Is this a table, schema, node, column, row, record? | Fact_Sales   | Discount      | Number          | %        | 50%      |                                    |   | Is this percentage discount or absolute price discount?   |
| Datasource2/ Products      | ProductID     | Is this a table, schema, node, column, row, record? | Fact_Sales   | ProductID     | Number          | #####    | 12345    | FK (references Dim_Products Table) | Sequence numbers are generated automatically. | Is this a varchar or a number?  |
| Datasource2/ Customers     | PostalCode    | Is this a table, schema, node, column, row, record? | Fact_Sales   | PostalCode    | Varchar(7)      | A#A #A#  | B3K 2R4  | FK (references Dim_Location Table) |   | Postal Code. Clarifying if this is for Canadian postal codes only, or for postal codes and zip codes for any country? For this exercise, the Canadian postal code will be used as the primary key for the target table. |
| Datasource2/ Customers     | CustomerID    | Is this a table, schema, node, column, row, record? | Fact_Sales   | CustomerID    | Number          | ####     | 1234     | FK (references Dim_Customer Table) | Sequence numbers are generated automatically. | Is this a number or varchar? Which of these two data sources are we using for CustomerID?   |
| Datasource1/ Customer      | CustomerID    | Is this a table, schema, node, column, row, record? | Fact_Sales   | CustomerID    | Number          | ####     | 1234     | FK (references Dim_Customer Table) | Sequence numbers are generated automatically. | Is this a number or varchar? Which of these two data sources are we using for CustomerID?   |
| Datasource2/ Orders        | OrderDate     | Is this a table, schema, node, column, row, record? | Fact_Sales   | Date          | Date            | MMDDYYYY | 12012017 | FK (references Dim_Time Table)     |   | Which of the to data sources are we using for date? Also need to identify format.   |
| Datasource1/ Book_Order    | OrderDate     | Is this a table, schema, node, column, row, record? | Fact_Sales   | Date          | Date            | MMDDYYYY | 12012017 | FK (references Dim_Time Table)     |   | Which of the to data sources are we using for date? Also need to identify format.   |

Source: <http://www.christoolivier.com/why-source-to-target-mapping-documents-matter/>

|                            |   |
|----------------------------|---|
| <b>Project</b>             | <i>Proposed Star Schema in a Data Warehouse</i>   |
| <b>ETL Package Name</b>    | <i>Load Dim_Location (US).dtsx</i>  |
| <b>Target Table</b>        | <i>Dim_Location_US</i>  |
| <b>Source Tables</b>       | <i>Customers</i>  |
| <b>Source Data</b>         | <i>Database to be loaded into Oracle</i>  |
| <b>Target DBMS</b>         | <i>Oracle</i>   |
| <b>Mapping Description</b> | Will create a target table Dim_Location_US to address US zip codes from source tables               |
| <b>Error Handling</b>      | On an error occuring any inserted records or updates needs to roll back and leave the system in the |

**Mapping Logic:**

| Source System/ Table  | Source Column | Source Type   | Target Table    | Target Column | Target Datatype | Format | Example | Constraint | Transformation/Business rule | Comments / Other rules                               |
|-----------------------|---------------|---|-----------------|---------------|-----------------|--------|---------|------------|------------------------------|--|
| Datasource1/ Customer | ZipCode       | Is this a table, schema, node, column, row, record? | Dim_Location_US | ZipCode       | char(5)         | #####  | 12345   | PK         |                              | Is this a U.S. zip code? Are we using U.S. zip code? |
| Datasource1/ Customer | City          | Is this a table, schema, node, column, row, record? | Dim_Location_US | City          | Varchar(25)     | ABCD   | Houston |            |                              | Is this a U.S. city?                                 |
| Datasource1/ Customer | State         | Is this a table, schema, node, column, row, record? | Dim_Location_US | State         | Varchar(25)     | ABCD   | Texas   |            |                              | Is this a U.S. state?                                |

## Orphan Source Data

1. We need to come up with plan for source data not accounted for in the target tables.
2. If we are not archiving the orphan source data, then we need a plan for staging into target tables.
3. In the meantime, we will store the following source tables and contents in an archive for easy retrieval when needed at a later date:

| Source System | Source Table     | Source Column     | Source Datatype | Format       | Example          | Constraint   | Comments / Other rules  |
|---------------|------------------|-------------------|-----------------|--------------|------------------|--|---|
| Data Source 1 | Author           | AuthorID          | Number          | ####         | 1234             | PK, FK (references Author from Author_Book)                              | Unique Author ID. Is this a number or varchar?  |
| Data Source 1 | Author           | AuthorName        | Varchar(50)     | Last,First   | Carandang, Carlo |  | Need to verify name format.<br>FirstName LastName? LastName, FirstName? Upper/ lower case?<br>Upper case first initial? |
| Data Source 1 | Book             | BookID            | Number          | ####         | 1234             | PK, FK (references Book from Author_Book, references Book from Ordering) | Unique BookID. Is this a number or varchar?   |
| Data Source 1 | Book             | Title             | Varchar(50)     | abcd         | Harry Potter     |  | Book title  |
| Data Source 1 | Book             | ISBN              | Number          | ####         | 1234             |  |   |
| Data Source 1 | Book             | Year              | Number          | ####         | 2000             |  | Year of publication   |
| Data Source 1 | Book             | Price             | Number          | \$           | \$10.45          |  | CAD? USD?   |
| Data Source 1 | Book             | NoPages           | Number          | ####         | 502              |  | number of pages   |
| Data Source 1 | Book             | BookDescription   | Varchar(100)    | abdc         | Reference Book   |  |   |
| Data Source 1 | Ordering         | Price             | Currency        | \$           | \$87.99          |  | CAD? USD?   |
| Data Source 2 | Employees        | EmployeeID        | Number          | ####         | 1234             | PK, FK (references Employees from Orders)                                | Unique EmployeeID. Is this a number or varchar?   |
| Data Source 2 | Employees        | FirstName         | Varchar(25)     | abcd         | Carlo            |  |   |
| Data Source 2 | Employees        | LastName          | Varchar(25)     | abcd         | Carlo            |  |   |
| Data Source 2 | Employees        | Title             | Varchar(50)     | abcd         | sales            |  |   |
| Data Source 2 | Employees        | WorkPhone         | Char(12)        | ###-###-#### | 902-123-4567     |  |   |
| Data Source 2 | Shipping_Methods | ShippingMethodID  | Number          | ####         | 1234             | PK, FK (references Shipping_Methods from Orders)                         | Unique ShippingMethodID. Is this a number or varchar?   |
| Data Source 2 | Shipping_Methods | ShippingMethod    | Varchar(50)     | abcd         | Air              |  |   |
| Data Source 2 | Customers        | CompanyName       | Varchar(50)     | abcd         | ADA              |  |   |
| Data Source 2 | Customers        | City              | Varchar(25)     | abcd         | Houston          |  |   |
| Data Source 2 | Customers        | State             | Varchar(25)     | abcd         | Texas            |  | Is this U.S. state?   |
| Data Source 2 | Customers        | PhoneNumber       | Char(12)        | ###-###-#### | 902-123-4567     |  |   |
| Data Source 2 | Payment_Method   | PaymentMethodID   | Number          | ####         | 1234             | PK, FK (references Payment_Method from Payment)                          | Unique PaymentMethodID. Is this a number or varchar?  |
| Data Source 2 | Payment_Method   | PaymentMethod     | Varchar(50)     | abcd         | Visa             |  |   |
| Data Source 2 | Payment          | PaymentID         | Number          | ####         | 1234             | PK   | Unique PaymentID. Is this a number or varchar?  |
| Data Source 2 | Payment          | PaymentAmount     | Number          | \$           | \$9.99           |  |   |
| Data Source 2 | Payment          | PaymentDate       | Date            | MMDDYYYY     | 10012016         |  |   |
| Data Source 2 | Payment          | CreditCardNumber  | Varchar(50)     | ####         | 1234             |  |   |
| Data Source 2 | Payment          | CreditCardEXPDate | Varchar(25)     | MMYY         | 1219             |  |   |
| Data Source 2 | Payment          | CardHoldersName   | Varchar(50)     | abcd         | Carlo Carandang  |  |   |
| Data Source 2 | Brands           | BrandID           | Number          | ####         | 1234             | PK, FK (references Brands from Products)                                 | Unique BrandID. Is this a number or varchar?  |
| Data Source 2 | Brands           | BrandDescription  | Varchar(50)     | abcd         | Lippincott       |  |   |