Whitney King

BUSIT 210 – Mod 2

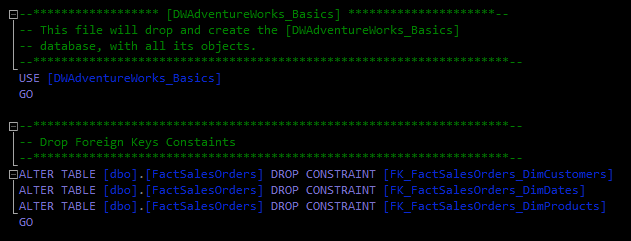
1/15/2017

*DWAdventureWorks\_Basics* ETL Script

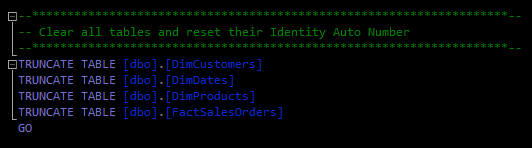
This ETL script is designed to extract important data from the AdventureWorks\_Basics database, and move it into a data warehouse structure that is better suited for end user consumption. It will first drop any data if the destination DB already exists, and then reload the tables with fresh information from the source DB. Additionally, it will create a date dimension to be used for looking up values for any date fields included from the source DB.

1. Drop foreign key constraint and truncate any existing tables
2. **ALTER TABLE** / **DROP CONSTRAINT** must be used on any existing foreign keys before you will be able to truncate the tables.

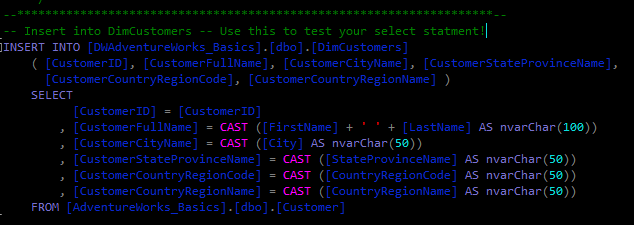
* If this data warehouse has been built before, there will be three existing foreign keys.



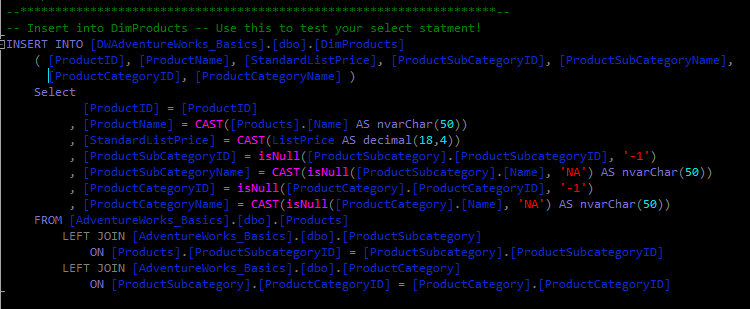
1. **TRUNCATE TABLE** will drop any existing information in the four tables being created and set the identity for the primary key columns back to 1.



1. Insert data from **Customer** into **DimCustomers**
   * Some fields will need to be transformed to the right number of characters for **nVarchar** types.
   * Additionally, customer **FirstName** and **LastName** should be combined into a single field.

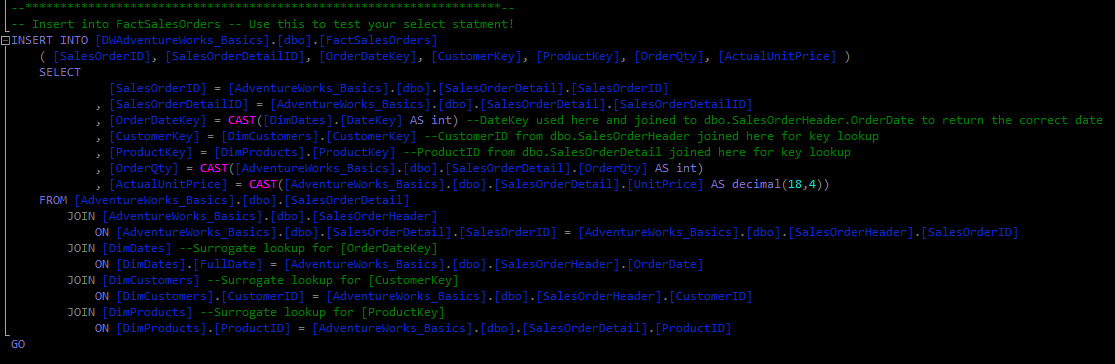
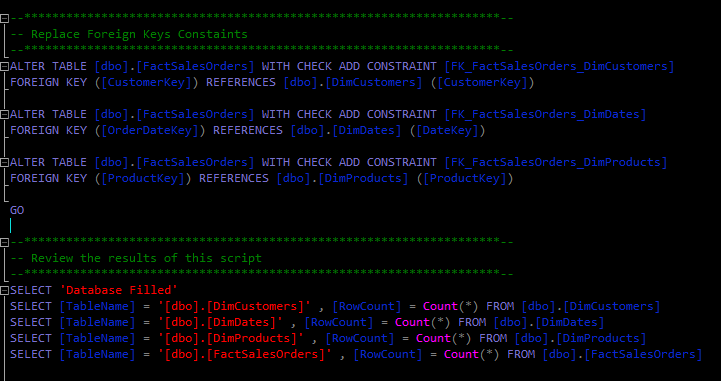


1. Join **Products**, **ProductSubcategory**, and **ProductCategory** tables and insert data into **DimProducts**
   * Data from three tables in the source DB will be joined to create one table in the destination DB.
   * To ensure there are no null values in the dimensions being filled, **isNull** is used to make sure any products without a category or subcategory are set to something.
   * **Money** values will be cast to **decimal** in the event the data needs to be handled in a tool that doesn’t support the money type, and other fields need to be cast to the specified number of characters for the **nVarchar** type.



1. Fill **DimDates** with values for the date range of 01/01/2005 – 01/01/2010
   * This data isn’t explicitly declared in the source DB so it needs to be created with a customized script that fills the necessary dates that may be needed when joined to date values in the various tables of the source DB.



1. Join data from the **SalesOrderDetail** and **SalesOrderHeader** tables in the source DB to surrogate keys from **DimDates**, **DimCustomers**, and **DimProducts** to fill the **FactSalesOrders** table.
   * **DimDates.FullDate**, **DimCustomer.CustomerID** **and DimProducts.ProductID** act as surrogate keys to link the fact table to each respective dimension, which will facilitate the creation of a foreign key constraint.
   * **OrderDateKey** and **OrderQty** should be cast to an **int** and **ActualUnitPrice** should be cast to a **decimal** value instead of **money**.
2. Replace the foreign keys that were removed at the beginning of the script, and output a **count** of how many rows were filled for each table where data is being inserted.
   * **ALTER TABLE** / **WITH CHECK ADD CONSTRAINT** is used for each foreign key that needs to be created, in this case one for each dimesion that has been filled with data and is connected to the fact table: **DimCustomers**, **DimProducts**, and **DimDates**.
   * The **CHECK** will ensure that the correct relationships exist between the keys in the tables for the data to be successfully joined.
   * The last **SELECT** statement will return a count of each table in the destination DB that has been filled with data – if there was an error during the processing of this ETL script, this part will not successfully execute to confirm the results.