Hand in 3

***Topic:***

ETL – scenario 2, S.C.D (Slowly changing dimension, initial load, and incremental load) process design and implementation for the AdventureWorks case. (Inspiration, reference case – NorthWind).

***Given case:***

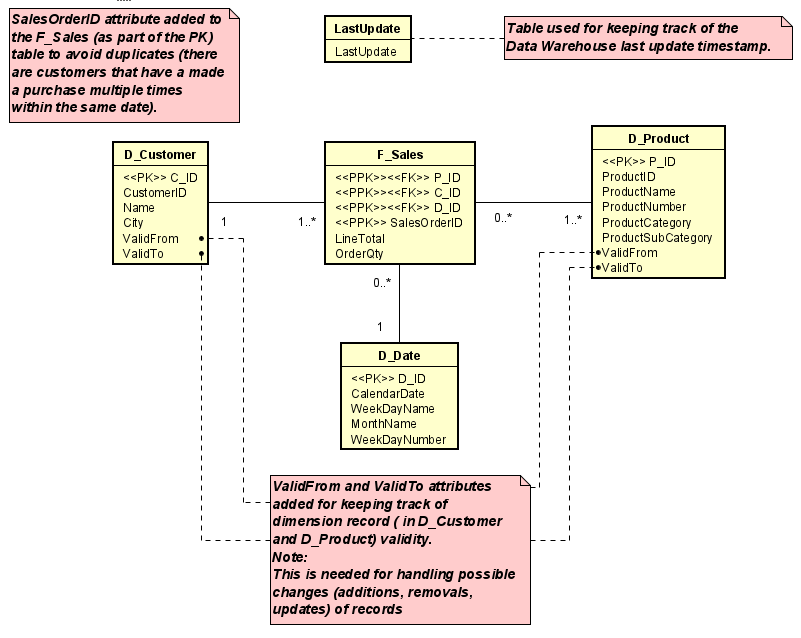
***Sales reporting****. Senior management would like to be able to track sales by customer and product, with the goal of establishing which products are the top sellers, which customers are the top buyers, and sales per month for top 3 selling products.*

*Note: Fact data split into 2 sets (online orders between 2011-05-31 and 2013-12-31 => initial load); (incremental load: 2014-01-01 and onwards)*

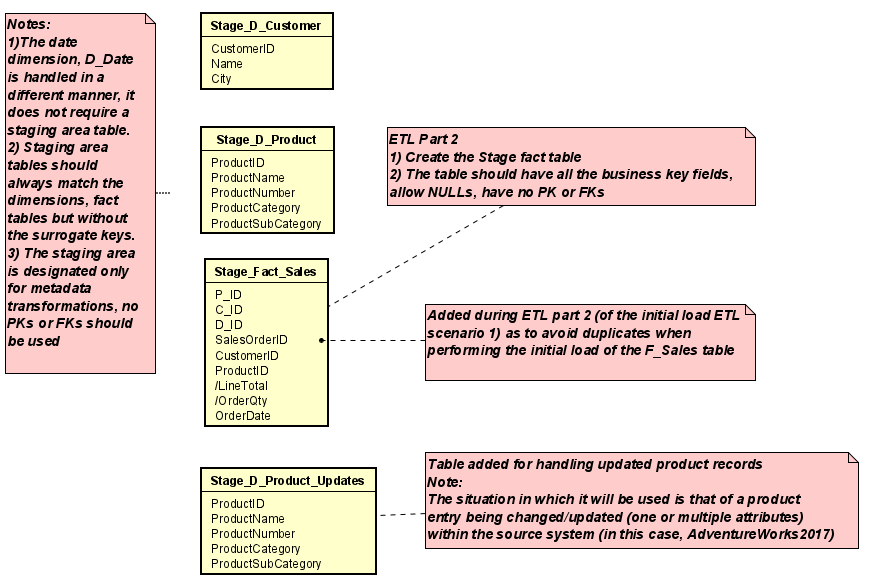
***Tasks:***

* *Document the updated version of the Data Warehouse through an ERD (include a table to handle the last updated date). Either the Product or Customer dimension should be able to handle slow changes (additions, deletions, updates of records within the source system).*
* *Document the initial load for dimensions and fact.*
* *Document the incremental load of data (including handling of SCDs for all cases, type 2 changes). This should be done for either the Product or Customer dimension.*

1. Updated Data Warehouse design (AdventureWorks\_DataWarehouse\_SCD)

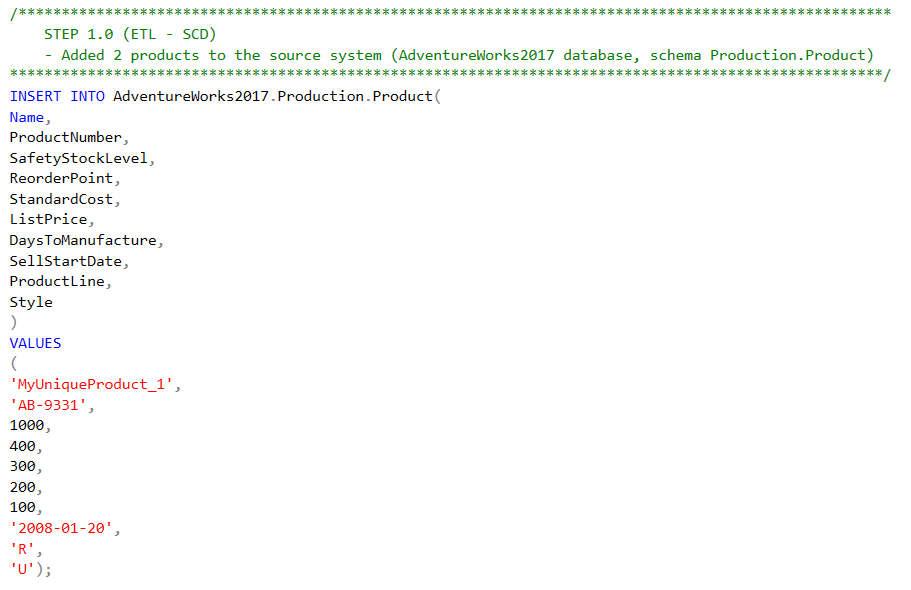


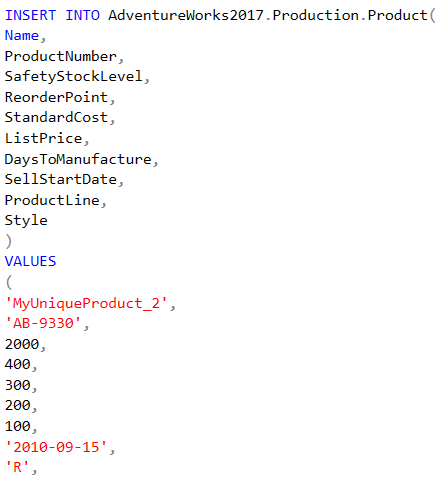
* 1. Updated Data Warehouse Staging Area design (AdventureWork\_StagingArea\_SCD)



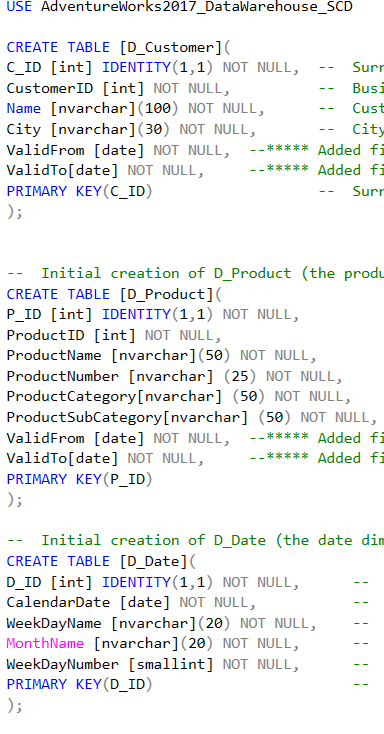
1. The ETL process design can be seen in the activity diagram ETL-AdventureWorks2017 SCD-Type 2
2. SQL Code for the initial load of dimension and fact data
   1. Adding two products to the source system (AdventureWorks2017)

These products will be used later for handling the removal and update cases of the slow changes within the D\_Product dimension.

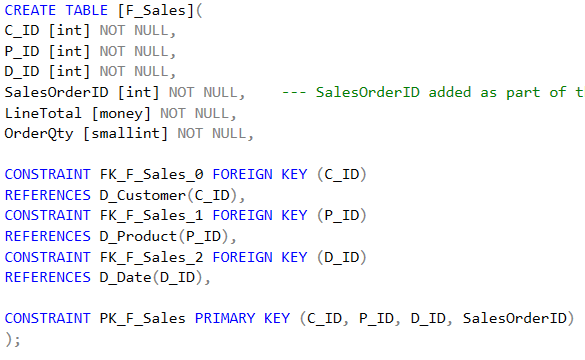




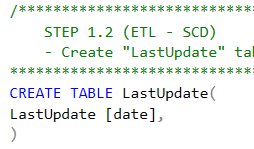
* 1. Creation of the Data Warehouse dimension tables



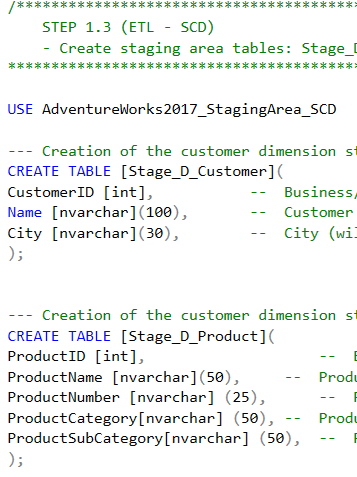
* 1. Creation of the Data Warehouse fact table



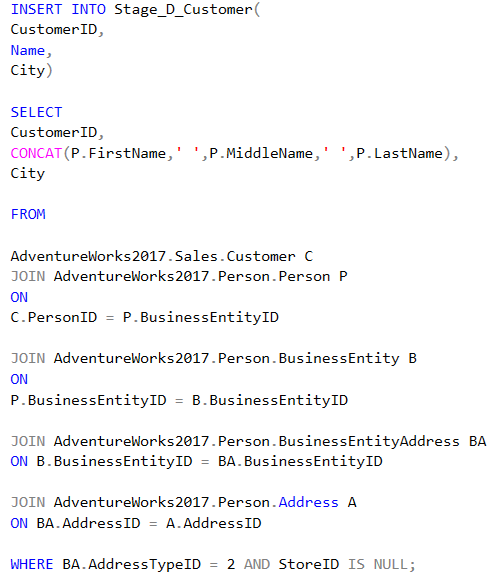
* 1. Creation of the Data Warehouse last update table

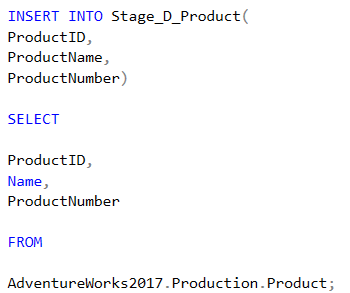


* 1. Creation of the Staging Area dimension tables (D\_Product and D\_Customer)

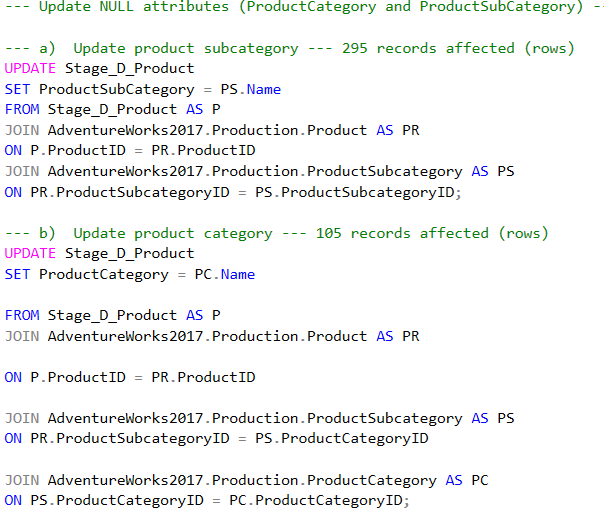


* 1. Populating the Staging Area tables (Stage\_D\_Customer and Stage\_D\_Product)

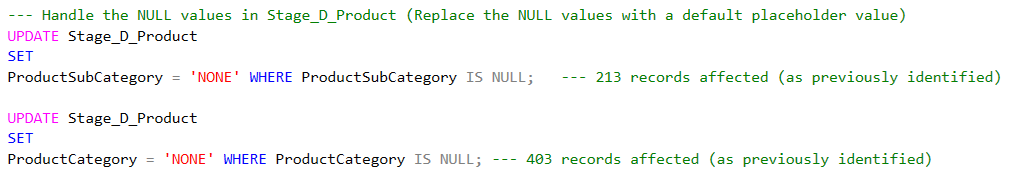




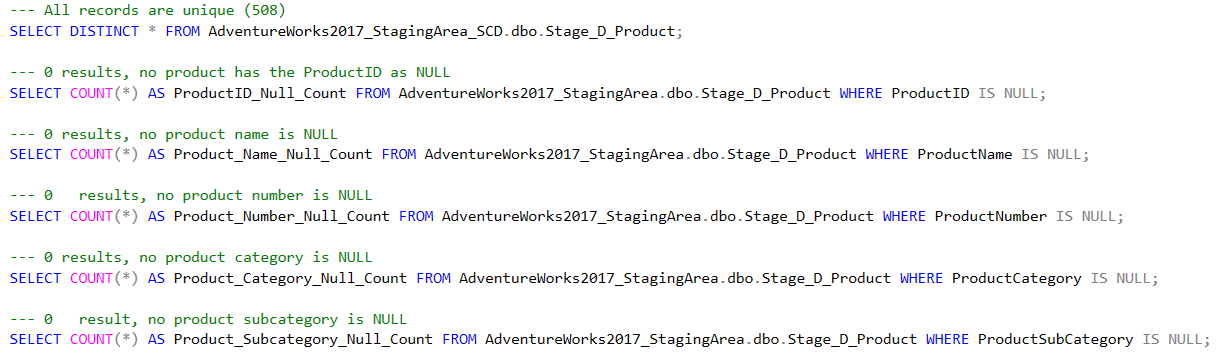
* 1. Transformations of Stage\_D\_Product data (ProductCategory and ProductSubCategory match)



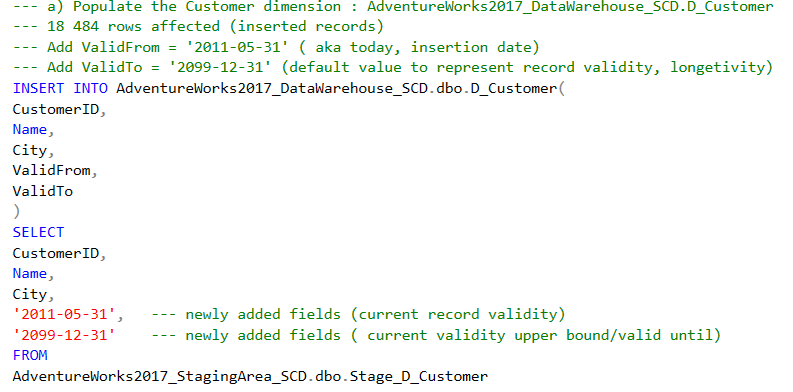
* 1. Transformations of Stage\_D\_Product data (Update of NULL values)



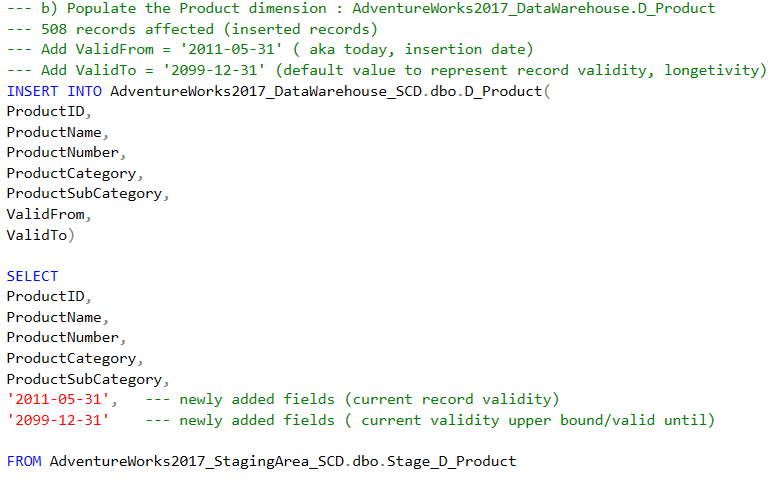
* 1. Verification of Stage\_D\_Product data



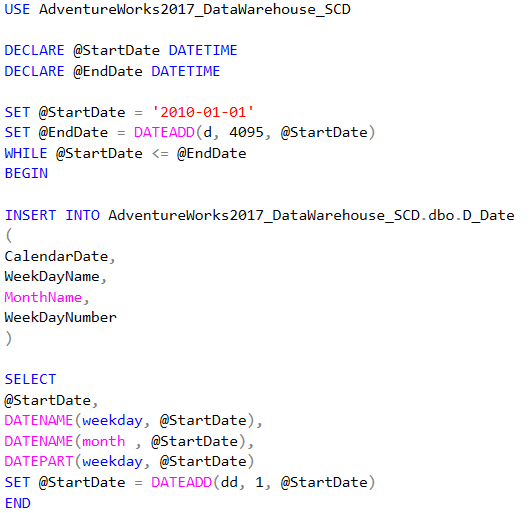
* 1. Populating D\_Customer – initial load (AdventureWorks2917\_DataWarehouse\_SCD)



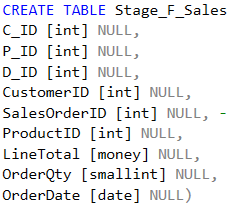
* 1. Populating D\_Product – initial load (AdventureWorks2917\_DataWarehouse\_SCD)



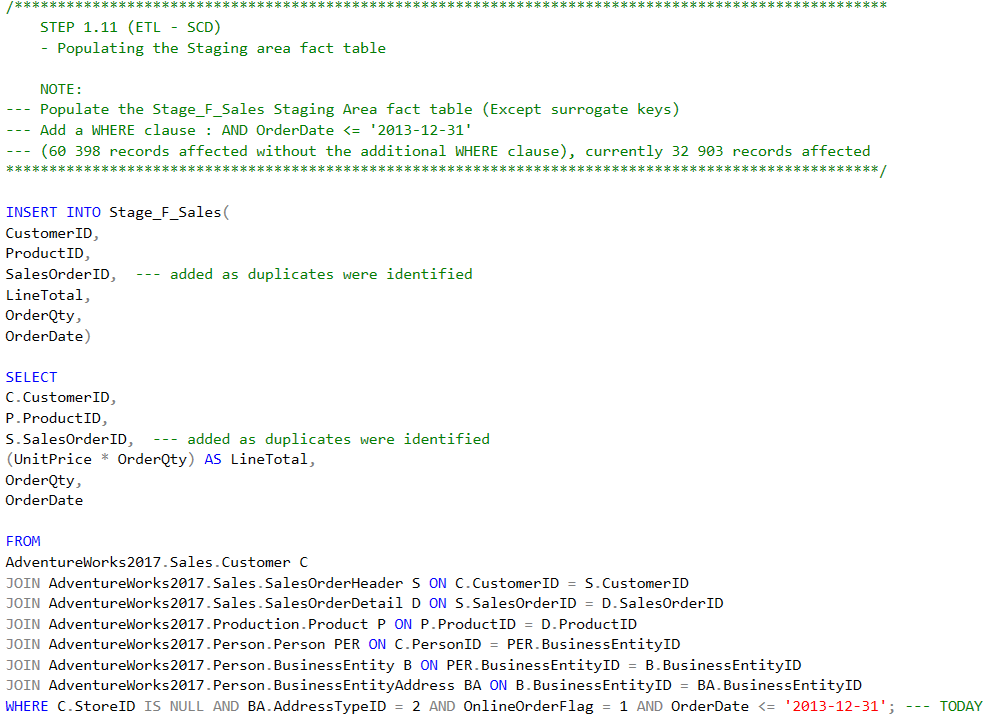
* 1. Populating D\_Date (AdventureWorks2917\_DataWarehouse\_SCD)



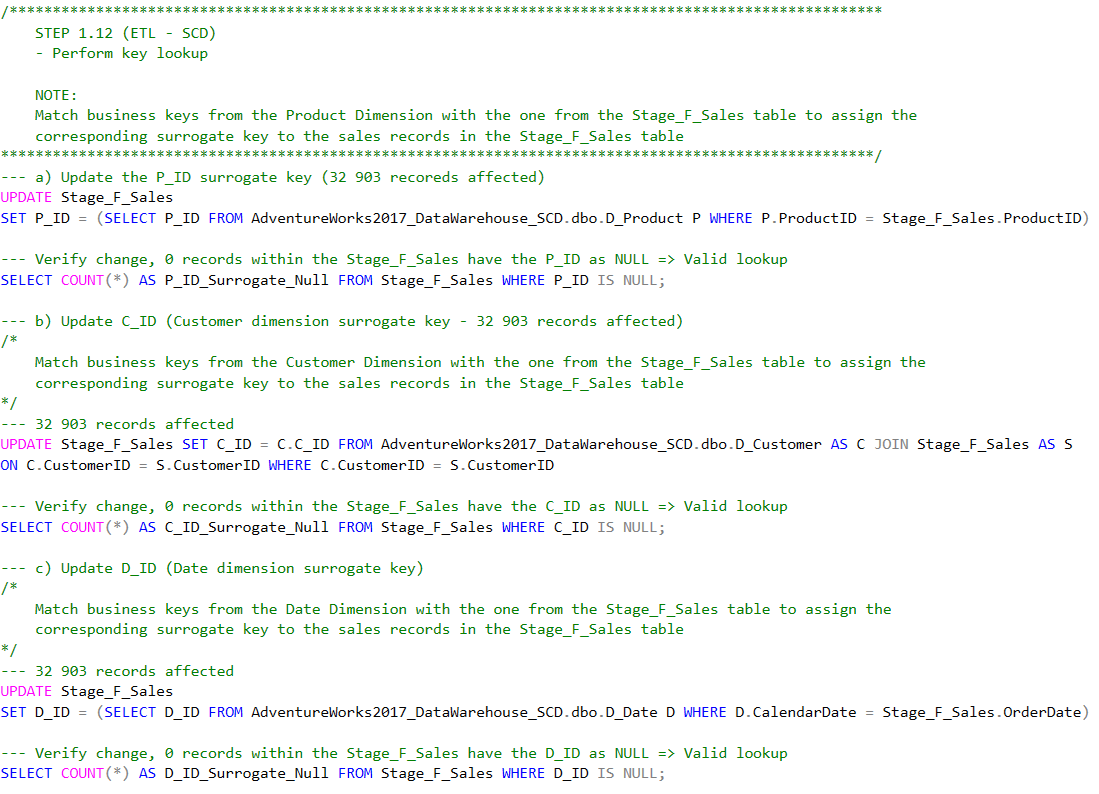
* 1. Creation of the Stage\_F\_Sales table (AdventureWorks2917\_StagingArea\_SCD)



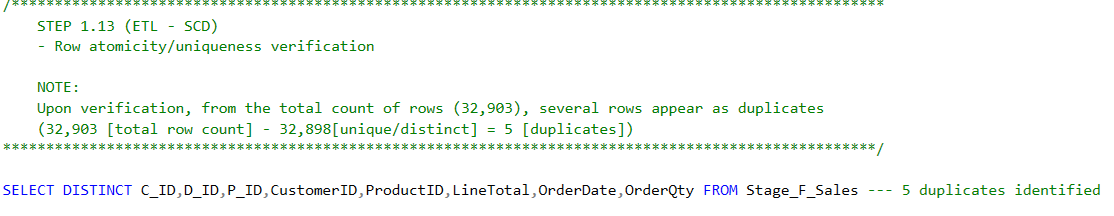
* 1. Populating the Stage\_F\_Sales table (AdventureWorks2917\_StagingArea\_SCD)



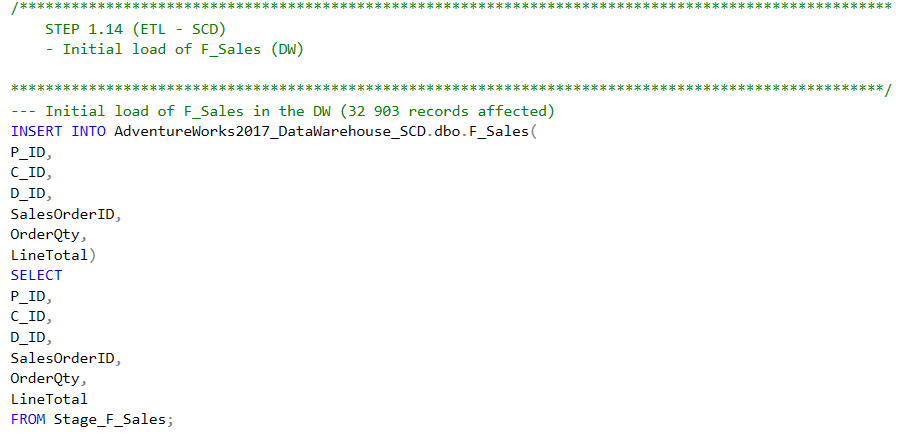
* 1. Performing the key lookup between the Stage\_F\_Sales table and the dimension (D\_Customer,D\_Product and D\_Date)



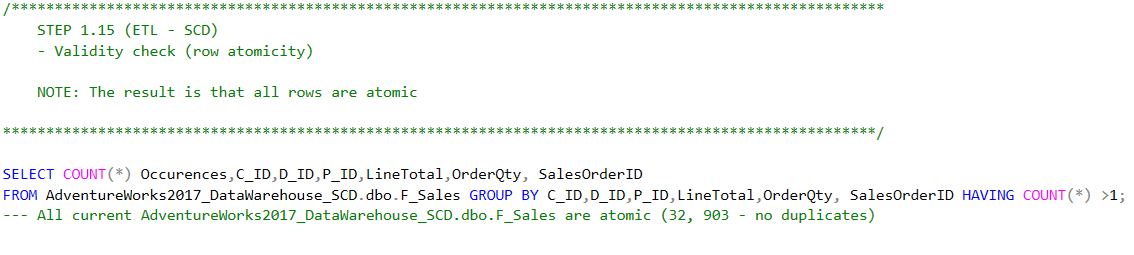
* 1. Row atomicity verification of the Stage\_F\_Sales table



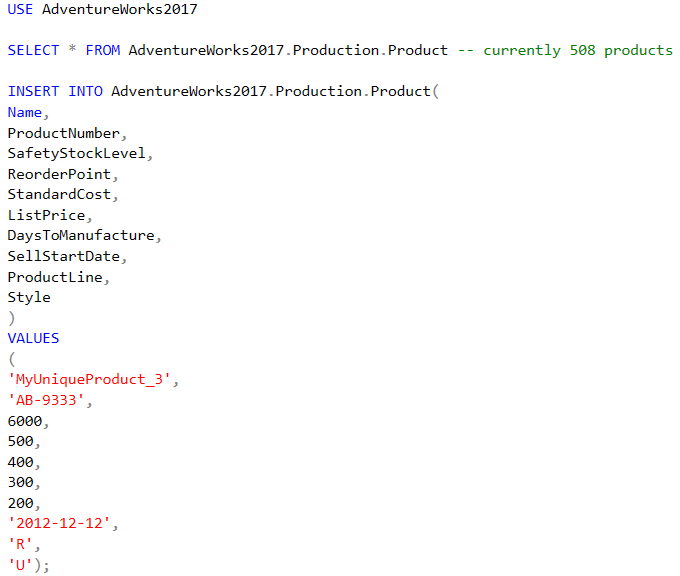
* 1. Initial load of F\_Sales of the AdventureWorks2017\_DataWarehouse\_SCD database



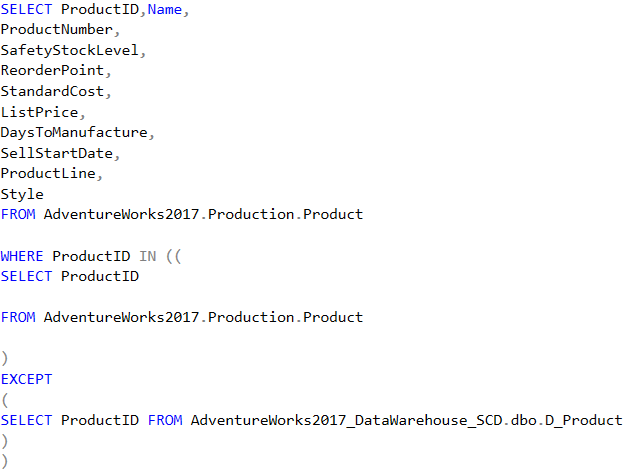
* 1. Validity check (row atomicity) for F\_Sales after initial load



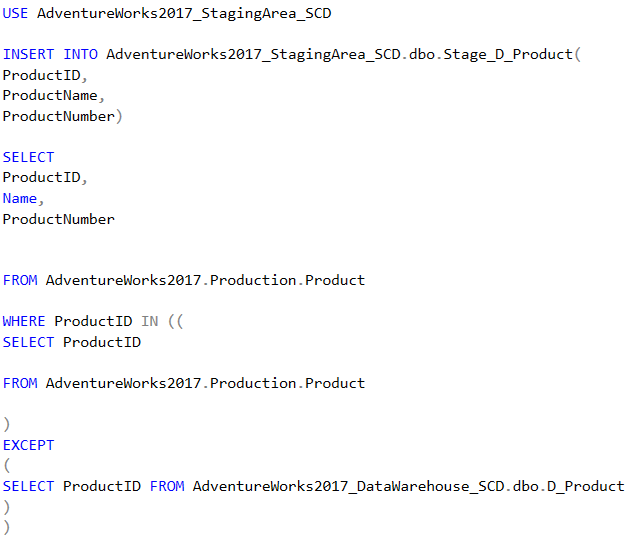
1. SQL Code for the incremental load of dimension and fact data (and handling of SCDs)
   1. Add a product to the source system



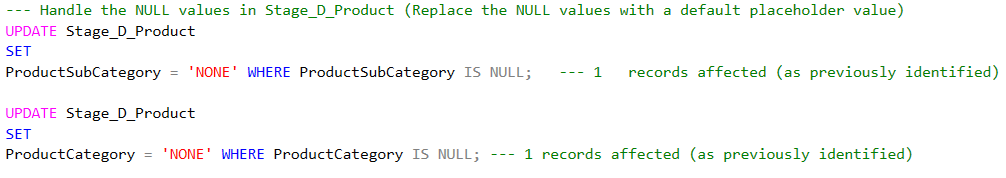
* 1. Finding the added product (performing set difference on ProductID between the Source System and D\_Product of the Data Warehouse)



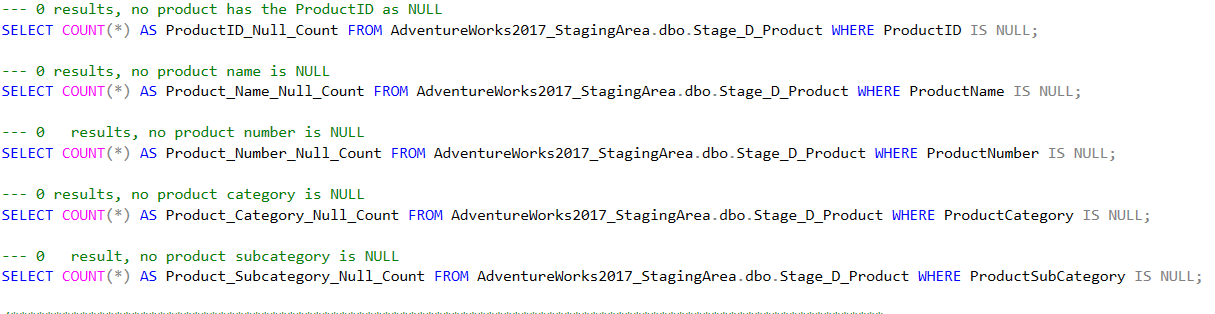
* 1. Adding the new product from the source system to the staging area



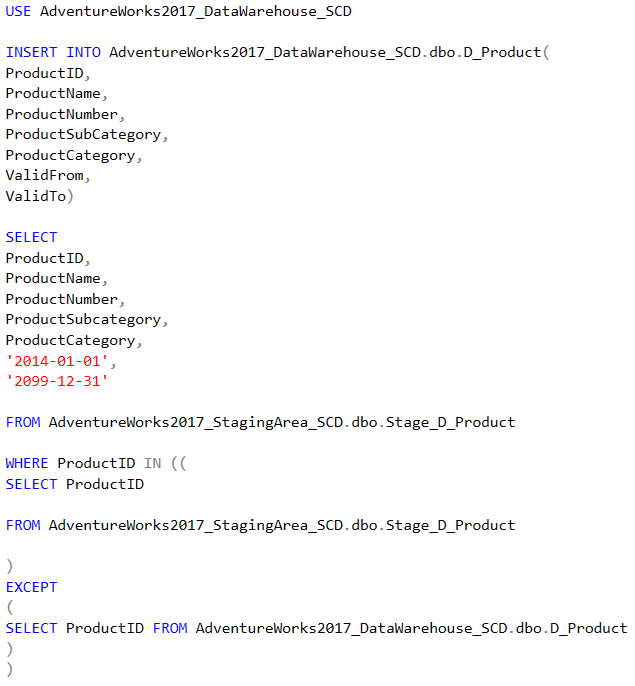
* 1. Perform transformations on the new product



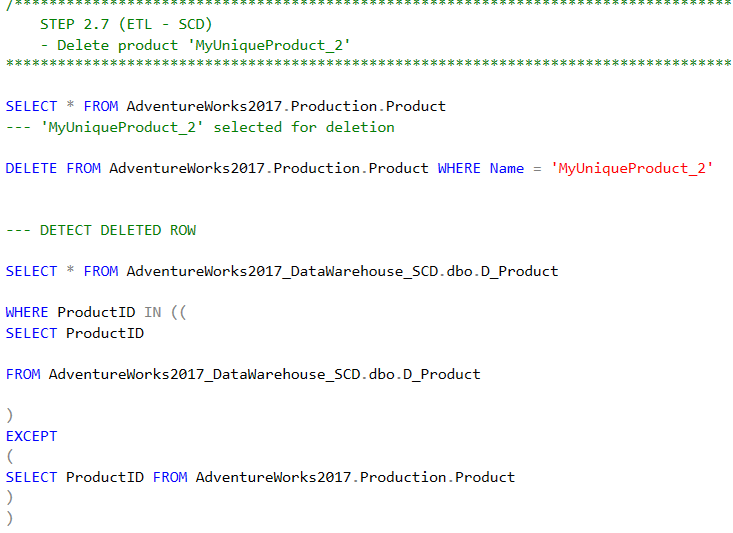
* 1. Perform verification of the updated record (attribute NULL check)



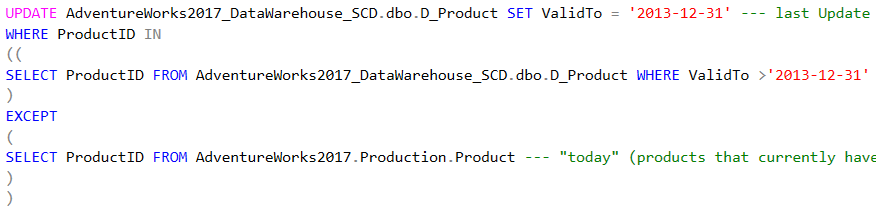
* 1. Load (Incremental) of the newly added product from the staging area to the D\_Product Data Warehouse table



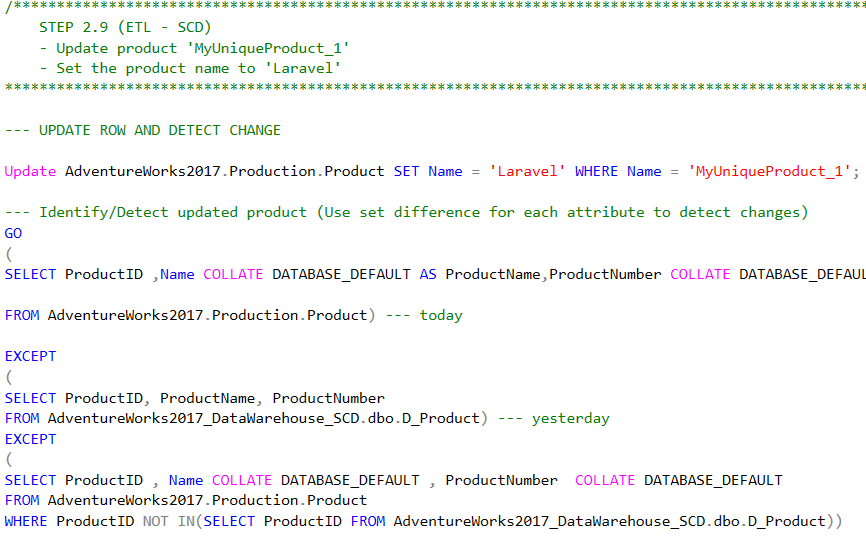
* 1. Deleting a product from the source system



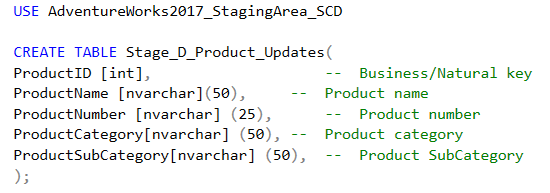
* 1. Update deleted product in the D\_Product dimension (change validity upper bound)



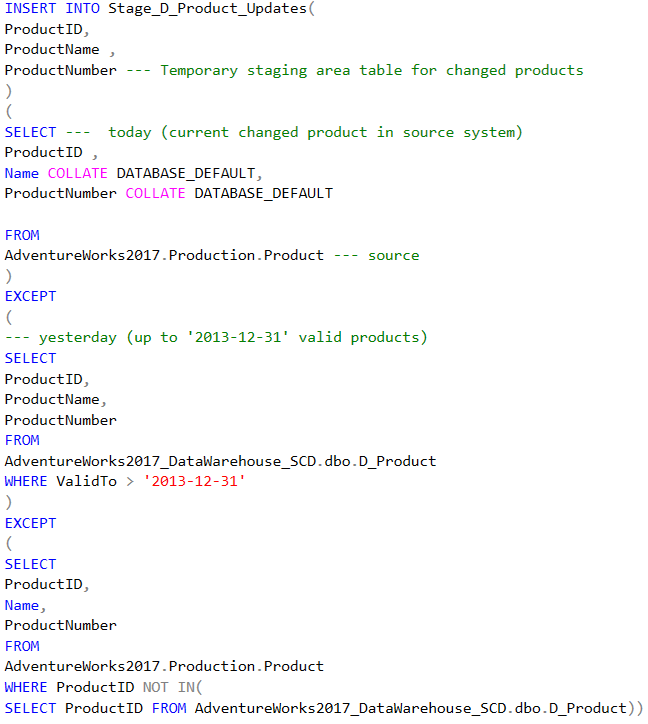
* 1. Updating a product (in the source system)



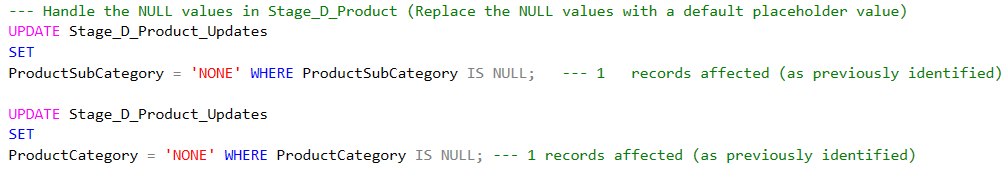
* 1. Creation of the Stage\_D\_Product\_Updates table (AdventureWorks\_StagingArea\_SCD)



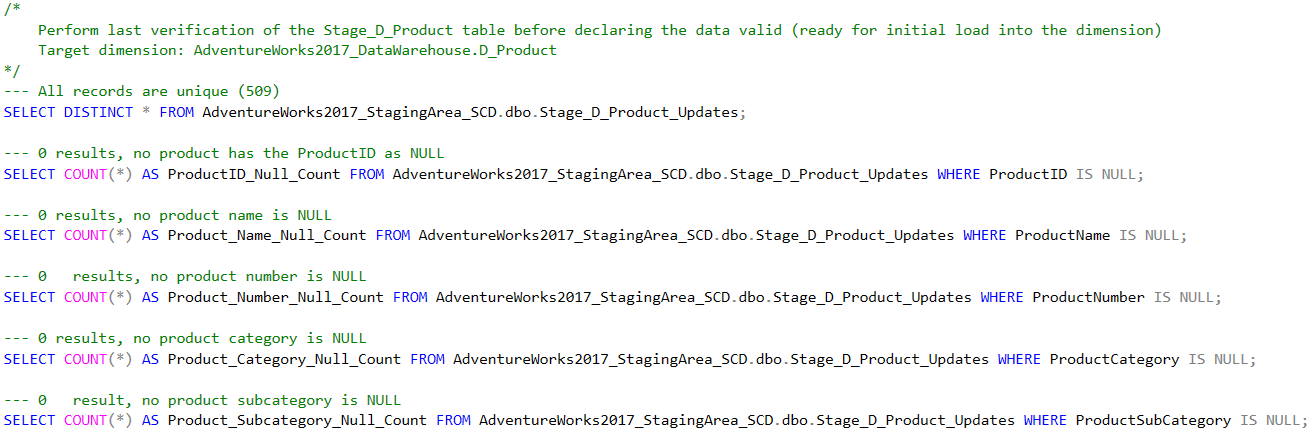
* 1. Inserting the updated record into the Stage\_D\_Product\_Updates table (by using set difference for record identification)



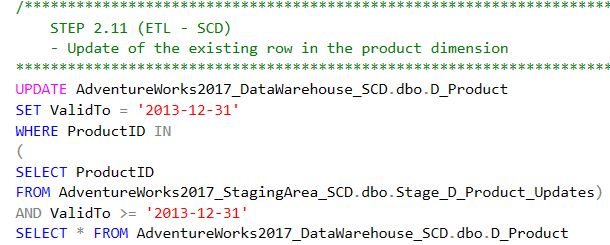
* 1. Transformations of the updated record in the Stage\_D\_Product\_Updates table



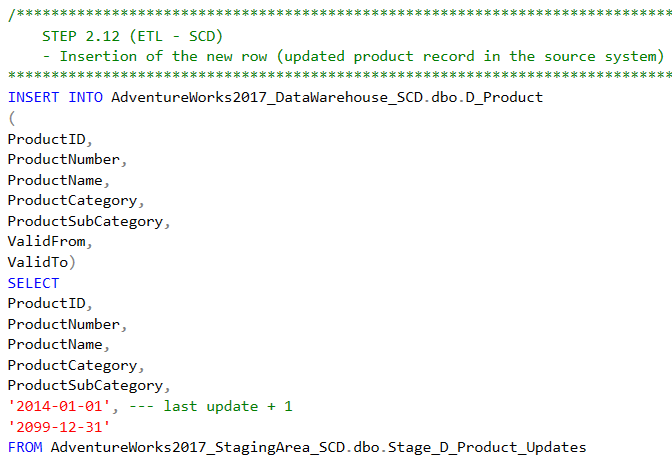
* 1. Last verification of the Stage\_D\_Product\_Updates table records (NULL check)



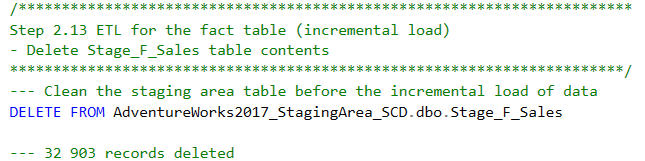
* 1. Update of the previously valid Dimension product record/entry – previously valid (matching by ProductID the updated product record of the source system)



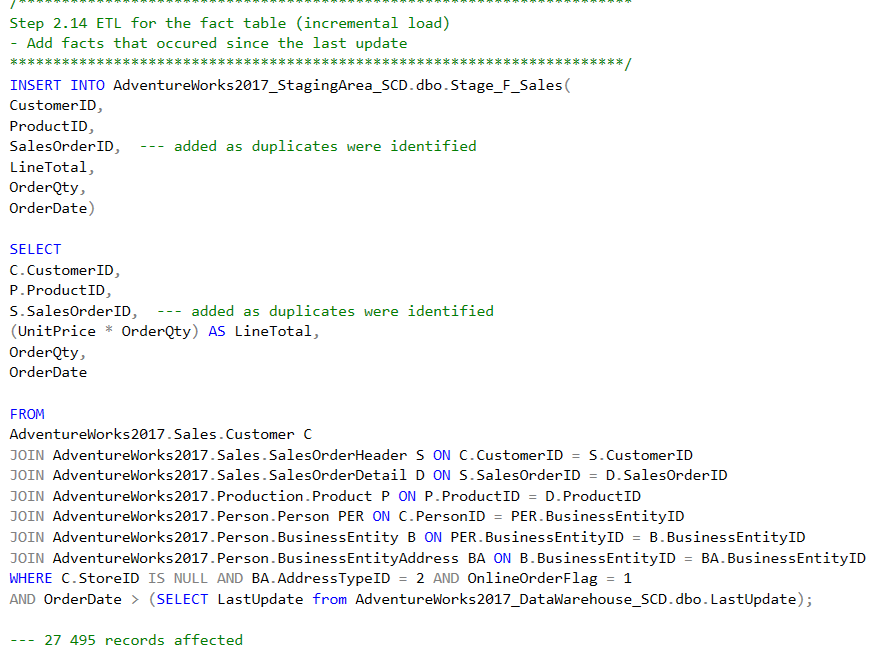
* 1. Load (Incremental) of the updated product record/entry – current valid (matching by ProductID the newly updated record of the source system)



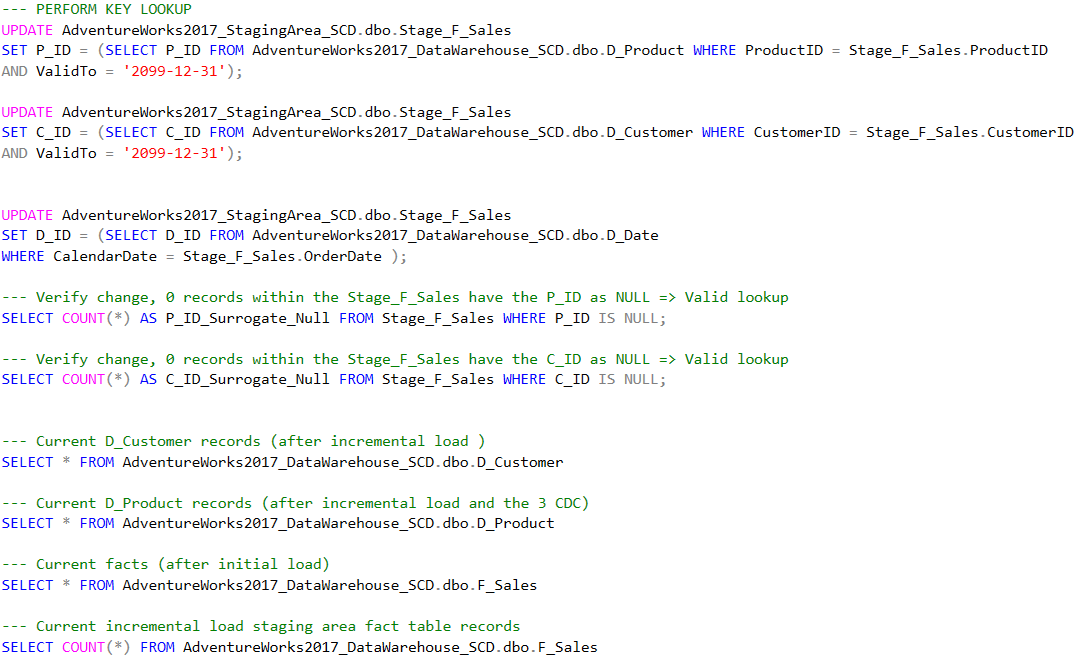
* 1. Content erase of Stage\_F\_Sales (preparation for the incremental fact record load)



* 1. Population of the Stage\_F\_Sales table with facts (sales) that occurred since the last update of the Data Warehouse (2013-12-31).



* 1. Key lookup and record verification for the Stage\_F\_Sales table facts (sales) that occurred since the last update of the Data Warehouse (2013-12-31).



* 1. Load (Incremental) of Stage\_F\_Sales table facts (sales) that occurred since the last update of the Data Warehouse (2013-12-31) into the Data Warehouse fact table (F\_Sales – AdventureWorks2017\_DataWarehouse\_SCD).

