**ETL Document :**

**Dimensional Model:**

**A computer screen shot of a computer

Description automatically generated**

**Cleaning & Transformation of Austin Datatset:**

A screenshot of a computer

Description automatically generated

1. Converting Crash\_date column to crash\_date\_actual by removing timeframe from it since it originally contained date as well as time.
2. Add distinct Record ID for each record before it is normalized.
3. Parse only the “Hour” from crash\_time and create a new integer datatype column to store it.
4. Check for the null values of all the columns and remove string nulls replacing it to “NA” and remove integer column nulls replacing it to “0”.

A screenshot of a computer

Description automatically generated

1. Concatenate contribution factor ID 1 and 2 columns using this expression in order to normalize them later.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

1. Use component Tnormalize to normalize the columns units\_involved and concatenated contribution\_factor\_id.
2. Map the related contribution\_factor with the values in the mapping document and extract the generalized values from the mapping document as a new column as we have to keep the id and the contribution factor description same for all the 3 datasets later.
3. Later apply upper case and trim functions to those column values which you will need to look up later.
4. Add columns DI\_CreatedDate, DI\_WorkflowName and DI\_SourceID to the table.

Complete Workflow Screenshot:

A screen shot of a computer

Description automatically generated

**Cleaning & Transformation of Chicago Datatset:**

A screenshot of a computer

Description automatically generated

1. Split crash\_date into crash\_Date\_string and crash\_time\_string by converting it into string first.
2. Add distinct Record ID for each record before it is normalized.
3. Check for the null values of all the columns and remove string nulls replacing it to “NA” and remove integer column nulls replacing it to “0”.

A screenshot of a computer

Description automatically generated

1. Concatenate primary\_contribution\_cause and secondary\_contribution\_cause columns in order to normalize them later.

A screenshot of a computer

Description automatically generated

1. Use component Tnormalize to normalize the concatenated column contributing\_cause.

A screenshot of a computer

Description automatically generated

1. Extract only the “Hour” from crash\_time\_string and create a new integer datatype column to store it.
2. Map the related contribution\_factor with the values in the mapping document and extract the generalized values from the mapping document as a new column as we have to keep the id and the contribution factor description same for all the 3 datasets later.
3. Later apply upper case and trim functions to those column values which you will need to look up later.
4. Add columns DI\_CreatedDate, DI\_WorkflowName and DI\_SourceID to the table.

Complete Workflow Screenshot:

A screenshot of a computer

Description automatically generated

**Cleaning & Transformation of NewYork Datatset:**

A screenshot of a computer

Description automatically generated

1. Add distinct Record ID for each record before it is normalized.
2. Check for the null values of all the columns and remove string nulls replacing it to “NA” and remove integer column nulls replacing it to “0”.

A screenshot of a computer

Description automatically generated

1. Concatenate all contributing\_factor\_vehicle and all vehicle\_type\_code columns in order to normalize them later.
2. Extract only the “Hour” from crash\_time\_string and create a new integer datatype column to store it.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

1. Use component Tnormalize to normalize the concatenated columns concat\_vehicle\_type and concat\_contri\_factor.
2. Map the related contribution\_factor with the values in the mapping document and extract the generalized values from the mapping document as a new column as we have to keep the id and the contribution factor description same for all the 3 datasets later.
3. Later apply upper case and trim functions to those column values which you will need to look up later.
4. Add columns DI\_CreatedDate, DI\_WorkflowName and DI\_SourceID to the table.

Complete Workflow Screenshot:

A screenshot of a computer

Description automatically generated

**Loading DimDate:**

A screenshot of a computer

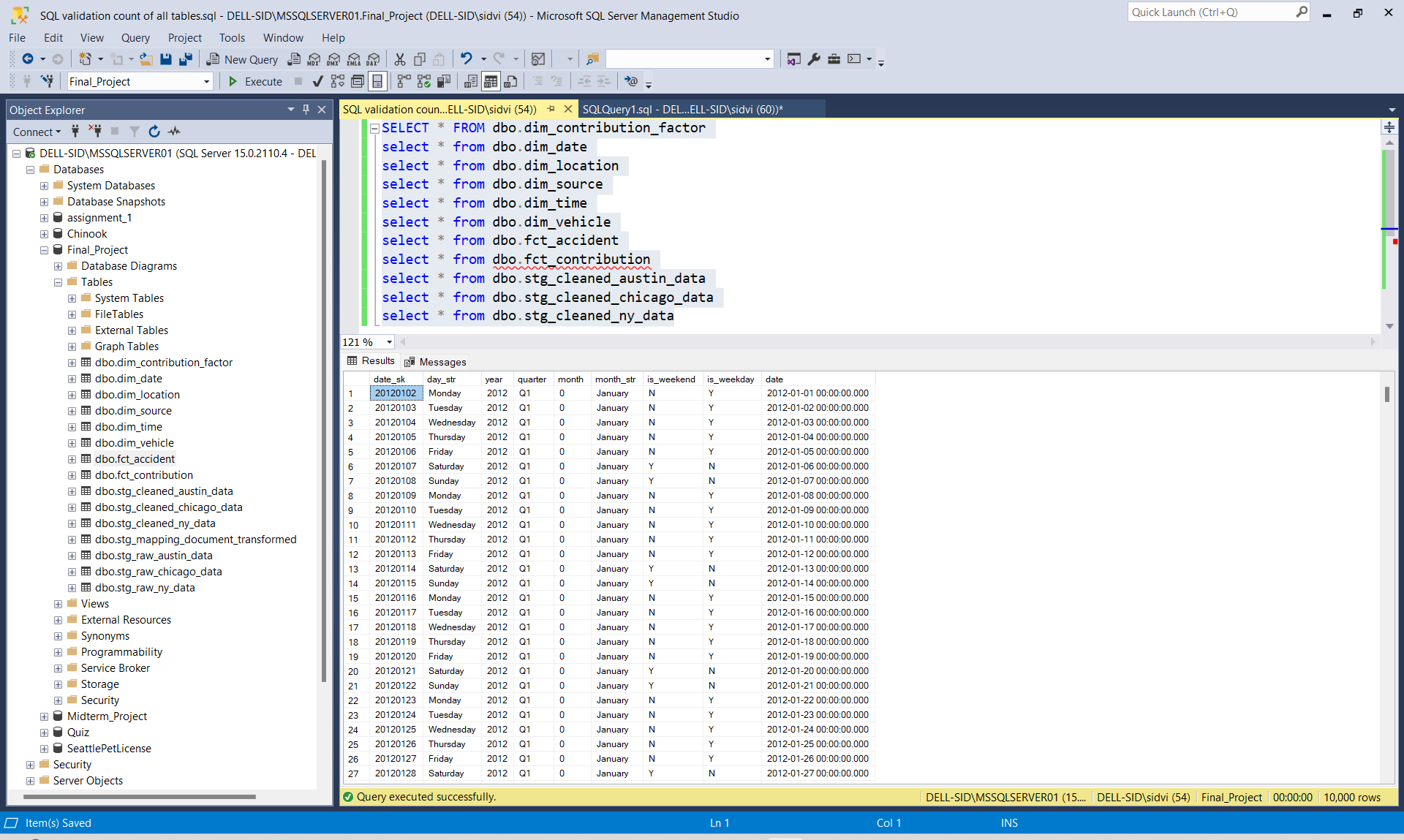
Description automatically generated

1. Use a TRow generator and keep the start date as 1 January 2012. Keep all the columns in hierarchy and also take the actual date value. Keep the holiday indicator as well. Add columns DI\_CreatedDate, DI\_WorkflowName and DI\_JobID to the table.

Complete Workflow Screenshot:

A screenshot of a computer

Description automatically generated



**Loading DimTime:**

A screenshot of a computer

Description automatically generated

1. Use a TRow generator to generate hours in integers from 0 to 23 and map the time period accordingly. Later we can lookup the data on “Hour” column in dim time and all the cleaned tables. Add columns DI\_CreatedDate, DI\_WorkflowName and DI\_JobID to the table.

Complete Workflow Screenshot:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Loading DimLocation:**

A screenshot of a computer

Description automatically generated

1. Select street names, latitude and longitude from all the cleaned tables. Unite them using a TUnite component and then extract only the unique set of values using the TUniqRow from all 3 combined. Add columns DI\_CreatedDate, DI\_WorkflowName and DI\_JobID to the table.

Complete Workflow Screenshot:

A diagram of a plane

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

**Loading DimVehicle:**

A screenshot of a computer

Description automatically generated

1. Select those columns from all the 3 cleaned tables which contain individual normalized records related to the vehicle involved in the accidents. Unite them using a TUnite component and then extract only the unique set of values using the TUniqRow. Add columns DI\_CreatedDate, DI\_WorkflowName and DI\_JobID to the table.

Complete Workflow Screenshot:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Loading DimSource:**

A screenshot of a computer

Description automatically generated

1. Select the column source name from all the 3 cleaned tables. Unite them using a TUnite component and then extract only the unique set of values using the TUniqRow. Add columns DI\_CreatedDate, DI\_WorkflowName and DI\_JobID to the table.

Complete Workflow Screenshot:

A map of a plane

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

**Loading DimContributionFactor:**

A screenshot of a computer

Description automatically generated

1. Select the column contribution\_factor and contrib\_factor\_id which are derived by mapping the code values and descriptions from the mapping document. Unite them using a TUnite component and then extract only the unique set of values using the TUniqRow. Add columns DI\_CreatedDate, DI\_WorkflowName and DI\_JobID to the table.
2. We have to implement SCD2 for this dimension so select output type as SCD2 table and do the following mappings.

A screenshot of a computer

Description automatically generated

Complete Workflow Screenshot:

A diagram of a plane

Description automatically generated

A screenshot of a computer

Description automatically generated

**Loading FctAccident:**

Loading fct\_accident using the cleaned Austin data:

A screenshot of a computer

Description automatically generated

1. Map the cleaned Austin data to a tmap. Take the lookups for all the 5 dimensions associated to the fact table. Join the columns rightly and apply a left outer join on all the unique matches. In order to calculate the measures, correctly analyze the data and implement the expressions by selecting the right attributes. Add a TUniqRow to select only the unique records(select comparison on all attributes to find a unique record.)

Complete Workflow Screenshot:

A computer screen shot of a network

Description automatically generated

Loading fct\_accident using the cleaned Chicago data:

A screenshot of a computer

Description automatically generated

1. Map the cleaned Chicago data to a tmap. Take the lookups for all the 5 dimensions associated to the fact table. Join the columns rightly and apply a left outer join on all the unique matches. In order to calculate the measures, correctly analyze the data and implement the expressions by selecting the right attributes. Add a TUniqRow to select only the unique records(select comparison on all attributes to find a unique record.)

Complete Workflow Screenshot:

A diagram of a network

Description automatically generated

Loading fct\_accident using the cleaned NY data:

A screenshot of a computer

Description automatically generated

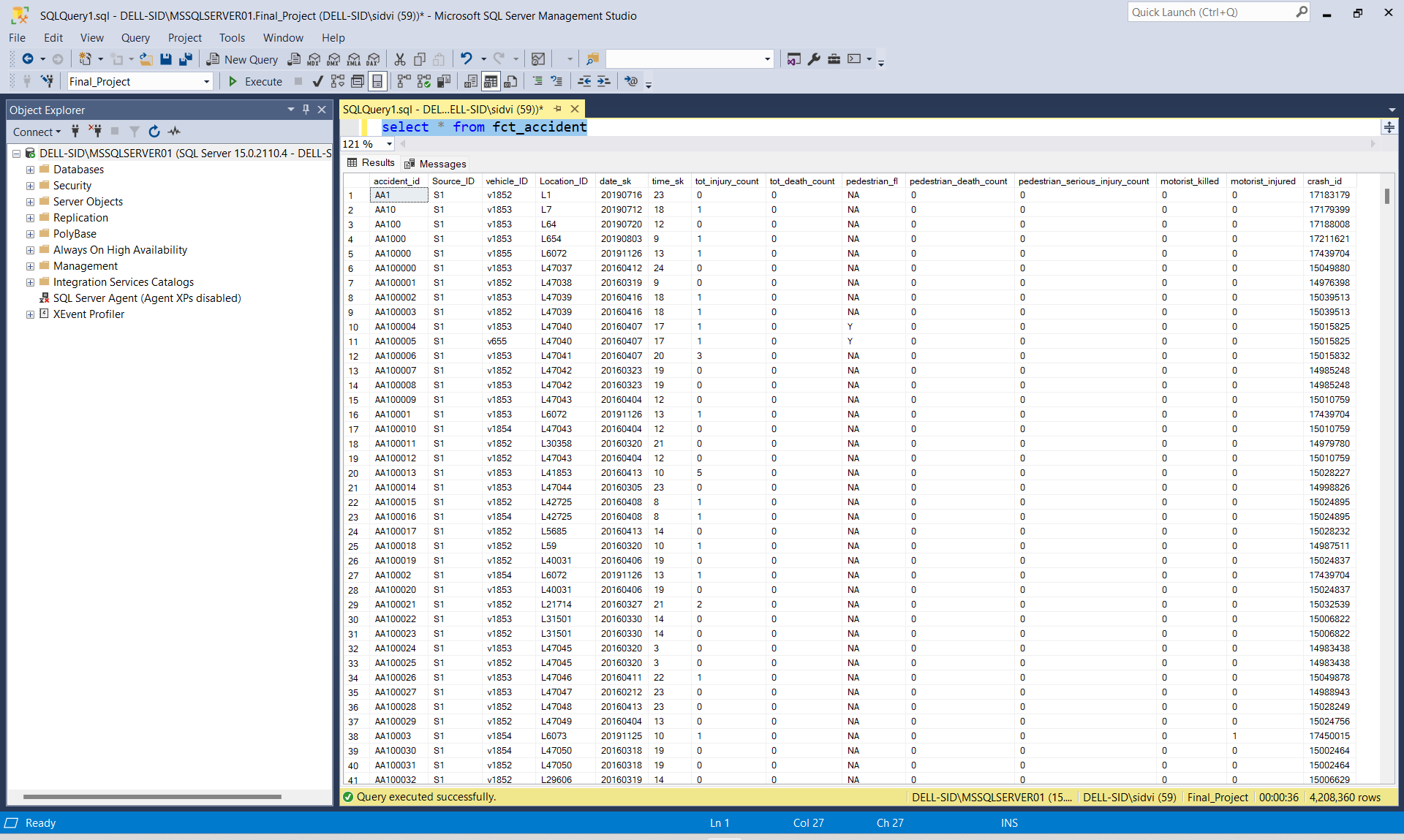
1. Map the cleaned NY data to a tmap. Take the lookups for all the 5 dimensions associated to the fact table. Join the columns rightly and apply a left outer join on all the unique matches. In order to calculate the measures, correctly analyze the data and implement the expressions by selecting the right attributes. Add a TUniqRow to select only the unique records(select comparison on all attributes to find a unique record.)

Complete Workflow Screenshot:

A computer screen shot of a network

Description automatically generated

**Preview of the main fact table : fct\_accident**



**Loading FctContribution:**

A screenshot of a computer

Description automatically generated

1. Map the cleaned data of all 3 tables to 3 distinct tmaps. Take the lookups for all the dimensions associated to the fact table. In this case, dim\_contribution will be the only one dimension. Lookup the main fact table fct\_accident as well. Join the columns rightly and apply a left outer join on all the unique matches.

Complete Workflow Screenshot:

A computer screen shot of a diagram

Description automatically generated

**Preview of fact table : fct\_contribution**

A screenshot of a computer

Description automatically generated