In order to move the right data set from source (NCDB_1999_to_2011.CSV) to destination (MS SQL Server 2012), we have organized our ETL (Extraction, Transformation and Loading) model into three layers:

1. DW_STAGING
2. DW_DATA
3. DW_NCDM

1) DW_STAGING LAYER:

The data present in the excel sheet is directly loaded into the Staging layer tables using the Import Wizard in the SQL Server Management Studio.
After loading all the data into master table (Tbl_Staging_NCDmaster), we have formed smaller tables out of it that corresponds to our dimensions and fact table (i.e. Table Collision, Date, Passenger, Vehicle, Fact).

Note: We have modified the Kimball's Date dimension according to our project requirements. For instance, we have added season column and removed some of the columns like day of month that are not appropriate for our requirements.

2) DW_DATA LAYER:

We have loaded the **cleaned** data in the Data layer, although, the overall structure of the data layer is identical to the staging layer tables. After extraction of the data, the main task is to transform (clean) the data that not only involves replacement of the values with the appropriate description present in the data dictionary apart but also looking for the NULL values and/or the duplicate values.
In order to do the same, we wrote different stored procedure for each of the dimension and fact tables that effectively replace all the values with given descriptions and checks for duplicate and null values. we have also created an error-logging table which helps us in keeping track of all the changes that we are making in our source database (i.e. Column name, table name, given value, changed value etc.).

Moreover, we have added a date timestamp column to all the tables in our database, which is acting as an **audit dimension** and this is very helpful when we want to perform incremental staging because by looking at a date timestamp value, user can know that at which particular date and time, we have loaded a particular set of data.

3) DW_NCDM LAYER (MIS LAYER):

The next layer is the NCDM layer (**final Data warehouse layer/Data Mart**) which consists of all the finalized cleaned, conformed dimensions and the conformed fact table.

The distinct feature of this layer as compared to the data layer is the **Lookup Tables** where we have created the **Surrogate Keys** for the given production keys in the database. These surrogate keys would be very useful to handle Slowly Changing Dimensions (Type 1, Type 2 and Type 3) and Production key Reuse.