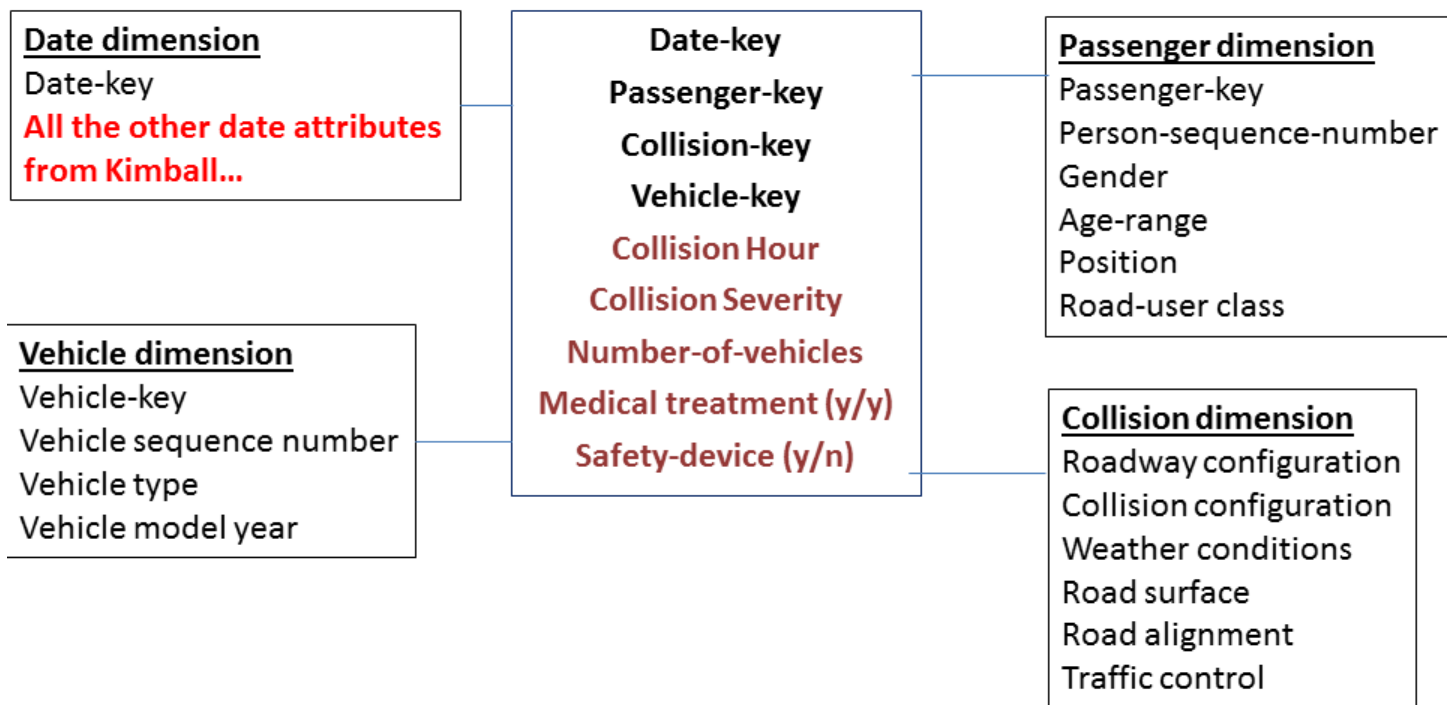


National Collisions Data Mart

Your task is to design and implement the National Collisions Data Mart from the data as obtained from Canada's Open Data repository.

Use all the data and metadata as located at <http://data.gc.ca/data/en/dataset/1eb9eba7-71d1-4b30-9fb1-30cbdab7e63a> to build your data mart. Your project will consist of data staging, together with the design of an OLAP end user interface, using any "full-fledged" DBMS, of your choice. Remember that you may use any programming language or tool to implement your interface.

Here is a dimensional model of the proposed data mart. You may use this model as is, or also modify it as you see fit.



Note that the use of the age-range attribute is **required** and that it should be created using attribute banding.

You will see that the CSV file uses codes. An important aspect of this project is to **transform** these codes into **meaningful descriptions**, to be used for decision support. That is, you are required to create textual, descriptive attributes in your dimensions. For example, for the collision configuration (C-CONF), you need to replace the numeric value (between 1 and 41) with a textual description. Similarly, you are required to replace the vehicle type codes (V-TYPE) with the textual descriptions of the vehicle type, the weather conditions codes (C-WTHR) to textual descriptions, and so on.

Others Notes:

1. Remember to create your own surrogate keys and full-fledged date dimensions.
2. You may add additional data if you wish, e.g. a weather dimension with more detailed weather information from Environment Canada. This is optional.

Refer to the “Typical Analytic Cycle” as described in class (and discussed in the text book) for a list of typical analytic questions which should be answered when accessing your data mart.

Here is a list of potential analytics we want to perform (note that this list is not complete and that I encourage you to also explore the data in other ways).

- a. Compare the total number of collisions in December 2013 with that of the previous ten years.
- b. Drill down from year to month to a specific day.
- c. Explore the data by type of accident.
- d. Explore the data by type of accident, given that a specific weather condition is present (e.g. drifting snow).
- e. Provide the details as to the severity of collisions, throughout the years.
- f. Compare the number of accidents by gender and/or age range.
- g. Explore the data in terms of the age ranges of fatalities.
- h. Explore the data in terms of the age ranges of fatalities, the type of vehicle, the season, and so on.
- i. Contrast the type of vehicles that are often in accidents, by accident type.
- j. Explore the impact of the road configuration, surface and/or road alignment on the frequency of accidents.
- k. Explore the typical number of vehicles involved in collisions, based on the day of the month.

Deliverables and Project Demonstrations

Submit your project using Blackboard Learn. Specifically, submit the following:

1. A zipped file containing all of your source code.
2. A copy of the following items:
 - a. A one (1) page high-level data staging plan, similar to Figure 10.1, on page 428, in the text book.
 - b. A one (1) page summary of your extract, transportation and load staging process. Explain how you loaded the dimensions as well as the fact table.
 - c. Your application template design, similar to Figure 12.3 on page 517 in the text book.
 - d. Your application navigation framework, similar to Figure 12.5 on page 520 in the text book.

You are required to demonstrate your project in a 15-20 minute timeslot, as scheduled by the Teaching Assistant.