University of Ottawa
School of Electrical Engineering and Computer Science
CSI4142 Project 2019

Ottawa Accidents Data Mart

This document contains the requirements for Phase 2 of the project.

Instructions

1. Complete this project in a group of (3) students.
2. Submit the project before the due date using the Virtual Campus.
3. Demonstrate your project to the corrector, in a 15 minute time slot, on March 13th, in SITE 4-010.
4. Use a database management system (DBMS) such as PostgreSQL to complete this project.

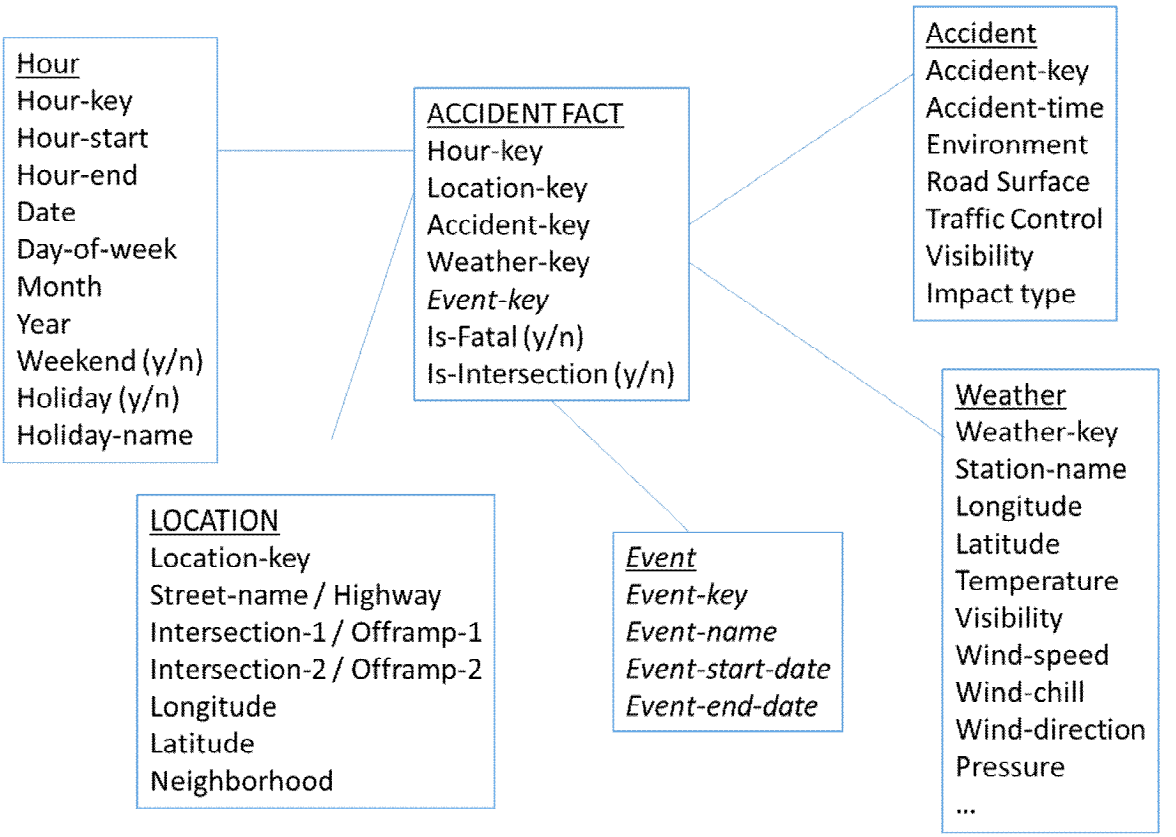
Deliverables:

Submit all your source code, together with a one-page high level data staging plan through the Virtual Campus**, before the due date**.
Note that all group members should submit the source code, not just one per group.
All group members should attend the project demonstrations.

Your task:

Your task is to design and implement an Ottawa Accidents Data Mart. Phase 2 of your project consists of data staging and writing OLAP queries.
You should use the data as provided at:
https://drive.google.com/drive/folders/1Mb67pKUWK9YpAK3I3FnQ0C81ewEdHn-x?usp=sharing
You are (optionally) welcome to augment it with additional enriched data, notably to contrast Ottawa with Toronto and Calgary. Note that the weather data includes all of Ontario and Alberta. You would need to extract the data for Ottawa.

Here is the dimensional model of the proposed data mart. You may use this model "as is", or modify and extend it as you see fit. Note that Is-Intersection is used a measure, in order to track the locations of accidents better. This attribute may also have been a descriptive attribute in the Accident dimension.



1. Remember to create your own surrogate keys. Refer to the slides and/or the book by Kimball et. al. that explain how to stage the data for surrogate key lookup.
2. You are encouraged to supplement the original data with enriched data from other sources, such as e.g. population statistics or other cities (e.g. Toronto and Calgary).
3. The data mart contains concept hierarchies on the Hour, Accident, and Location dimensions.
4. Event is an optional dimension that tracks events such as Hockey games, Winterlude and Fall Rhapsody.
5. Refer to the "Typical Analytic Cycle" as described in class, for a list of typical analytic questions which should be answered when accessing your data mart.

Requirements and Mark Allocation

Note that Phase 2 is out of **70**. Additional work, outside the scope of this project, could earn you up to 20 additional marks.

1. **(10 marks) Physical Design:** Create the physical schema of the data mart using the DBMS of your choice.

2. **(30 marks) Data staging**: Extract and transform the data and load all rows into the data mart. Be sure to record all the steps that you followed and submit a one-page high level schematic with your source code.
   During this phase, you should assess the quality of the data and show how you determined that the road surface, environment and visibility indicators in the accident dimension correspond to the nearest weather station data.

3. **(30 marks) OLAP queries**: Implement the following five types of queries by traversing the concept hierarchies. (Refer to Example 4.4 on page. 146 of the Data Mining textbook (3rd edition) by Han et. al. for a description of these operations.)
   a. **Drill down.**
   b. **Roll up.**
   c. **Slice.**
   d. **Dice.**
   e. Top N or Bottom N (Iceberg queries).

   These operations enable us to answer questions such as the following. (Note that these are just illustrative examples, and that your system should be designed to answer these categories of queries.)

   1. **Drill down, roll up**: Determine X in { *accidents, fatal accidents, not fatal accidents* } per Y in { *year, month, day, hour, location (street, intersection, neighborhood* }
      For instance, determine the total number of fatalities in Ottawa during the four years.
      For instance, determine the total number of fatalities in Ottawa during 2015.
      For instance, determine the total number of fatalities during an ice storm in Ottawa during 2014.
      For instance, determine the total number of fatalities during an ice storm in Ottawa during December 2014.
      For instance, determine the total number of fatalities during an ice storm in Downtown Ottawa during 2014.
      For instance, determine the total number of fatalities during an ice storm in Downtown Ottawa during December 2014.

2. **Drill down, roll up, slice and dice**: Contrast X in {*accidents, fatal accidents, not fatal accidents*} per Y in {*year, month, day, hour, location (street, intersection, neighborhood*} when compared to Z in {*year, month, day, hour, location (street, intersection, neighborhood)*}
   For instance, compare the number of accidents on Mondays, versus the number of accidents on Fridays.
   For instance, compare the number of fatal accidents on Mondays, versus the number of fatal accidents on Fridays.
   For instance, contrast the number of fatal accidents in Nepean, during 2017, with the number of fatalities in Orleans, during 2014.
   For instance, contrast the number of fatal accidents in Nepean on Mondays, with the number of fatalities in Nepean on Fridays.
   For instance, contrast the number of fatal accidents in Nepean on Mondays between 05h00 and 08h00, with the number of fatalities in Gloucester during the same period of time.

   For instance, contrast the number of accidents at intersections versus those that do not occur at intersections.
   For instance, contrast the number of fatal accidents at intersections versus those that do not occur at intersections.
   For instance, contrast the number of fatal accidents at intersections in Downtown Ottawa versus those that do not occur at intersections.

3. Explore the interplay between weather conditions and accidents, focussing on traffic control, intersection, impact type and visibility.
   For instance, contrast the total number of accidents during summer, with the number of fatalities during fall.
   For instance, determine the interplay between road surface and the number of accidents.
   For instance, determine the interplay between prolonged heat waves with high humidity levels and the number of accidents.
   For instance, determine the interplay between strong rain showers and the number of accidents.
   For instance, contrast the number of accidents in Nepean during summer, with the number of fatalities in Orleans during winter.
   For instance, contrast the number of accidents in Nepean when it snows, with the number of fatalities in Nepean when it rains.
   For instance, determine the interplay between traffic control, impact types and frequencies of accidents.

4. Locate "hot spots" for certain types of accidents, focussing on traffic control, intersection, impact type and visibility.
   For instance, determine the intersections with the most accidents over the four years.
   For instance, determine the neighborhoods with the most accidents during evening rush hour.
   For instance, determine the intersections with the most accidents in fall, during dusk.
   For instance, determine the sections of highways with the most accidents when visibility is poor.

5. Calculate trends over the years.
   For instance, determine the weekly trends in fatal accidents in Downtown Ottawa during the four years.
   For instance, determine the monthly trends in adverse weather conditions over the four years.