**University of Ottawa**

**School of Electrical Engineering and Computer Science CSI4142 Fundamentals of Data Science**

**Project Phase 1: Conceptual Design – Dimensional Model**

**Data Science Team 12:**

Badr, Elhattab Elibrahimi

Ange Michaella, Niyonkuru

Aisha, Khalid

## Data Mart Grain

An individual test case with results and with the weather, location trends, holiday, testing facility, and restrictions for that day.

## Dimensions & Dimensional Attributes

### Age Bands Dimension

* 1. **age\_band\_id**: *integer, PK*
  2. **Min\_age**: *integer*, minimum = 0 and maximum = 200, sample value = 40
  3. **max\_age**: *integer,* minimum = 10 and maximum = 210, sample value = 50

### Person Dimension

* 1. **person\_id**: integer, PK
  2. **age\_band\_id**: integer, FK
  3. **gender**: char, [‘F’, ‘M’, ‘U’] *(female, male, unspecified)*

### Date Dimension

* 1. **date\_id**: integer, PK
  2. **month\_name**: string, sample = ‘December’
  3. **month\_num**: *integer*, minimum = 1 and maximum = 12, sample value = 4
  4. **day\_name**: *string*, sample = ‘Monday’
  5. **day\_num**: *integer*, minimum = 1 and maximum = 7, sample value = 4
  6. **date**: *date, sample: 2021-02-02 (format YYYY-MM-DD)*
  7. **year**: *integer*, minimum = 1900 and maximum = 3000, sample value = 2021
  8. **week\_of\_year**: integer, minimum = 1 and maximum = 53, sample value = 4
  9. **week\_of\_month**: integer, minimum = 1 and maximum = 5, sample value = 3

### Testing Facility Dimension

* 1. **testing\_facilty\_id**: integer, PK
  2. **test\_capacity**: integer, minimum = 1 and maximum = MAXINT, sample value = 4
  3. **hours\_open**: integer, minimum = 1 and maximum = 24, sample value = 4
  4. **date**: date, *sample: 2021-02-02 (format YYYY-MM-DD)*

### Restriction Dimension

* 1. **restriction\_id**: integer, PK
  2. **intervention\_category**: string, sample: travel, distancing
  3. **intervention\_type**: string, sample: work from home, isolation
  4. **description**: string, sample: Mandatory use of face coverings for all air and rail passengers announced
  5. **organization**: string, sample: Canada, Ontario
  6. **start\_date**: date, *sample: 2021-02-02 (format YYYY-MM-DD)*

### Weather Dimension

* 1. **weather\_id**: integer, PK
  2. **mean\_temp\_C**: integer, minimum = - 100C and maximum = 100C, sample value = 4
  3. **precipitation\_mm**: integer, minimum = 0mm and maximum = 50000mm, sample value = 4
  4. **precipitation\_type**: string, sample: rain, snow
  5. **date**: date, *sample: 2021-02-02 (format YYYY-MM-DD)*

### Holiday Dimension

* 1. **holiday\_id**: integer, PK
  2. **month**: *integer*, minimum = 1 and maximum = 12, sample value = 4
  3. **day**: *string*, sample = ‘Monday’
  4. **name**: *string, sample:*
  5. **statutory**: *boolean, TRUE or FALSE*

### Location Trends Dimension

* 1. **Loc\_id**: *integer, PK*
  2. **Percent\_change**: *integer,* minimum = - 5000 and maximum = 5000, sample value = 4
  3. **Type**: *string, sample: grocery\_and\_pharmacy, retail\_and\_recreation*
  4. **Province**: *string, sample: Ontario*
  5. **District**: *string, sample: Ottawa Division, Toronto Division*

## Measures & Facts

Detail all the measures/facts. You should list the domains and sample values.  
(*e.g., Age: integer, minimum = 0 and maximum = 130, Sample value = 35*).

### Measures

* + 1. **Status**: *string, sample: fatal, resolved*
    2. **Acquisition**: *string, sample: travel, CC, No known EPI link, unspecified EPI link*

### Facts

* + 1. **Date\_id**, FPK
    2. **Holiday\_id**, FPK
    3. **Loc\_id**, FPK
    4. **Weather\_id**, FPK
    5. **Restriction\_id**, FPK
    6. **Person\_id**, FPK
    7. **Test\_facility\_id**, FPK

## Assumptions Made

* Only general (not individual) restrictions are being considered.
* Daily aggregates are to be calculated using data in the fact table.

## Work Plan

Our group virtually met using voice chat and screen sharing to collaborate. We had a total of 3 collaborative meetings on:

1. January 30th (1 - 2:30 pm):
   * We looked at various datasets together, created a rough draft of the conceptual model based on our initial look of the data and questions we needed to answer. We also chose which datasets we would use.
2. January 31st and (4 - 5:30 pm):
   * We looked at datasets more indepth to find specific attributes and make sure we had data for all the attributes. Created a more finalised version of the dimensions and attributes. We were not sure if we should have a second fact table, we decided to email our professor and meet again to work more on the report and the dimensional model.
3. February 2nd (4:30 - pm)
   * Decided on having one fact table for individual cases and went over the attributes again to finalize. We also spent time writing the report, and making sure the dimensional model had all the data needed to answer the different questions mentioned in the project deliverable instructions.

## 6. Diagram

