**Introduction**

Considering the provided Relational Schema derived from ECDC, it is consisted of 8 tables (Virus, Pandemic, Location, Time, Vaccine, Vaccination, Target\_Group, ICU).

Primary Keys are underlined. Foreign Key are denoted as Attribute name 'Table Name'

1. **Data Warehouse Model**

Data warehouse modelling helps in designing the schemas with more details and summarized information of any data warehouse. The aim of data warehouse modelling is to create a schema presenting the reality of the fact in data warehouse.

VIRUS (V-ID, Name, Family, Type, Infection\_rate)

PANDEMIC (L-ID: Location, T-ID: Time, V-ID: Virus, Current\_Cases, New\_Cases, Recovered, Deaths

LOCATION (L-ID, Country, Region, Population

Time (T-ID, Year, Month, Date, Start Date)

Vaccine (Vac-ID, Name)

Vaccination (Vac-ID: Vaccine, T-ID: Time, L-ID: Location, Trg-ID: Target\_Group, Doses\_Recieved, First\_Dose, Second\_Dose, Third\_Dose

Target\_Group (Trg-ID, Age\_Range)

ICU (L-ID: Location, T-ID: Time, Indicator, Source, ICU\_Cases)

* 1. **Attribute Tree from Rational Schema**

Considering the provided ECDC relational model, all the attributes, relationships and values have been identified.

Below figure 1 is the result after connecting all of the entities.

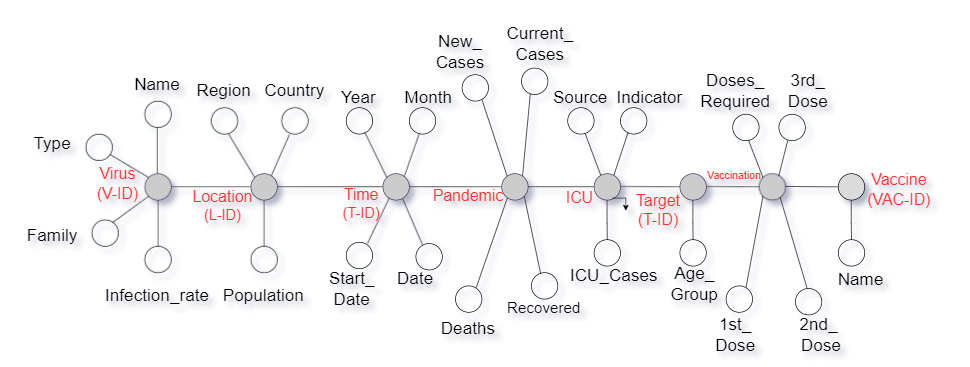


Figure 1

* 1. **Pruning of Attribute tree**

Here the designed data is compressed in order to reduce the size of attribute tree. Non critical and redundant sections have been removed. Below figure 2 reflects the pruned design as compare to figure 1.

Pruning is also done by taking a consideration note that the information gain at any particular node is greater than minimum gain.

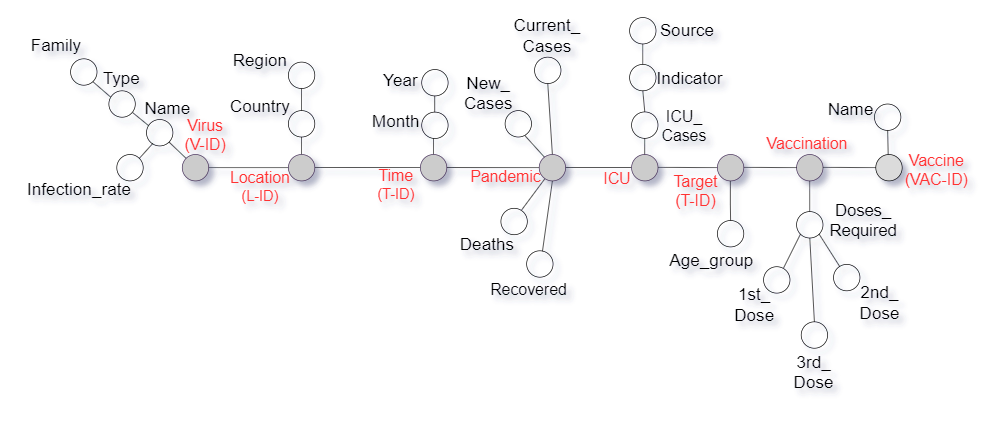


Figure 2

* 1. **Fact Schema from Attribute Tree**

An ad-hoc and graphical representation is done. This is supporting the conceptual modelling phase by providing the information in terms od Dimension, Fact and Measures.

* + 1. **Identification of Attributes like Dimension, Fact and Measures**

The below figure 3 identifies that there are 3 major fact and 5 dimensions.

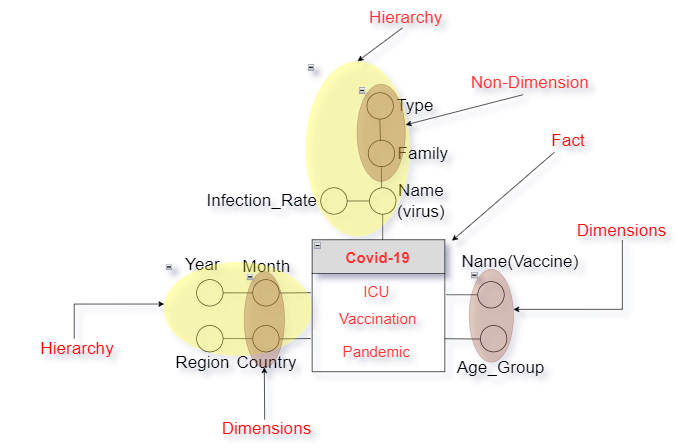


Figure 3

* + 1. **Correctness of Fact Schema**

The figure 3 gives the reflection of correct number of facts, dimension and fact attribute. The hierarchies are set after pruning the attribute tree. All the unwanted attributes have been removed without losing any information.

1. **Logical Model**

A graphical connection has been shown in this report which helps in understanding the nature of the obtained warehouse.

* 1. **Mapping of DFM model to a Logical Model**

In this report, primary keys have been specified with respect to all the entities. Relationship between the entities has been described and have found all the attributes for each entity. Normalization of the data is done in order to present the connection.

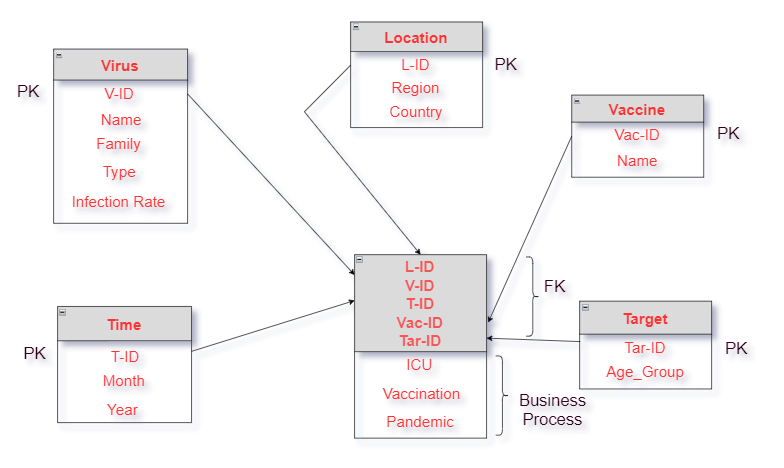


Figure 4

* 1. **Classification for Fact Table and Dimension**

Above Logical Model figure 4 show the connection between PK and FK. All the PK are listed as FK in Fact table. Basis the PK IDs, information on business process can be obtained.

* Fact box of the fact schema leads to create a fact table.
* The dimension leads to create a dimension table.
* First dimension attribute of each dimension become foreign key
* Measure attribute becomes fact table attribute

­­

1. **Data Warehouse Schema**

Under this Schema, report has presented a logical explanation of the whole structure. It includes the name and explanation of records. Above figure 4 explains all connections between the attributes. Using Data definition language (DDL) statements, structure of database has been created.

The DDL commands in SQL create database schema to define the structure of the data that will be stored in a database. Either Export the Dataset or The CREATE query is used to create a database or objects such as tables, views etc.

In this report, Database has been created using SQL queries. The following example demonstrates the CREATE query for MySQL Server:

>> CREATE DATABASE Covid19 The script above creates a database named “Covid19” in MySQL Server. The CREATE query has also been used to add tables in an existing database as shown in the following script:

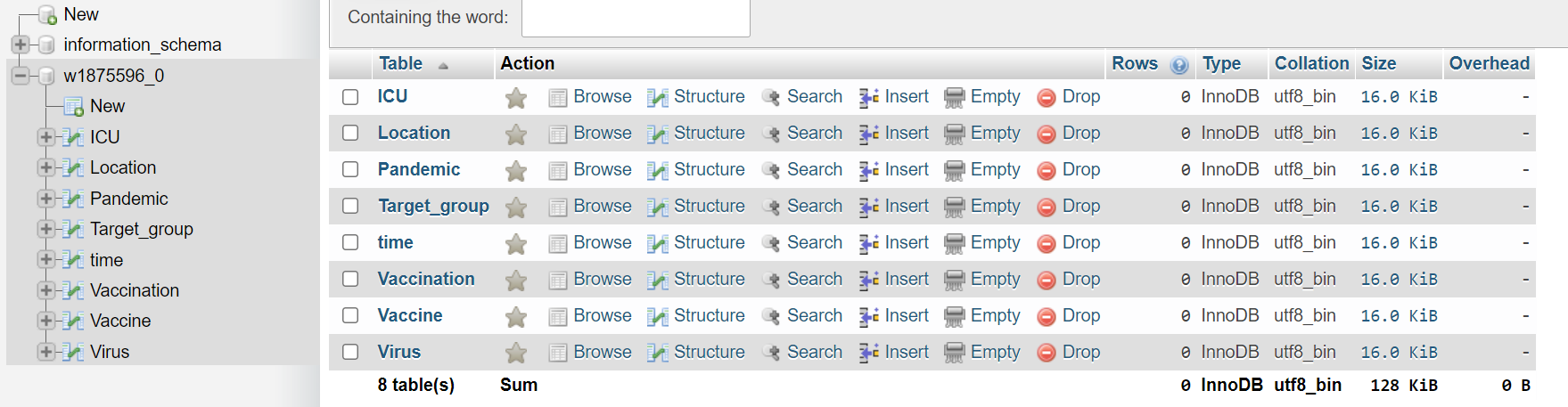


Figure 5

The DDL Queries can be found in the attached file –



1. **Introduction to Materialized View**

Using Materialized View, the data is modified from the view definition query and is programmed to update the data changes in the primary tables. This improves the reading of complex queries.

CREATE MATERIALIZED VIEW is used to create a materialized view. The FROM clause of the query can be the name tables, views & other materialized views.

* 1. **Calculation of Fact Table Size**

Size of any fact table can be calculated roughly using the below formula: (referenced)

* Small: 1 KB/row x 1000 rows = 1 MB.
* Medium: 5 KB/row x 100k rows = 500 MB.
* Large: 10 KB/row x 1m rows = 10 GB.
  1. **Matrix specification**

Data Warehouse matrix comprises the names of Facts on rows and the names of Dimensions in the columns The associations among the Facts and Dimensions is the value in the grey cell. A “ ” icon in the cell at the node of a Fact row and Dimension column specifies that the Fact is related to that dimension.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Facts | Vaccine | Age\_Group | Month | Region | Name | Dimension Usage Count |
| ICU |  |  |  |  |  | 2 |
| Vaccination |  |  |  |  |  | 4 |
| Pandemic |  |  |  |  |  | 3 |
| Fact Usage Count | 1 | 1 | 3 | 3 | 1 |  |

Table 1

* 1. **Justification of Materialized View**

Materialized views respond faster and improve the performance of queries. Materialized views are repeatedly and transparently sustained by Snowflake. After making any changes with the database, a background service keeps materialize view updated.

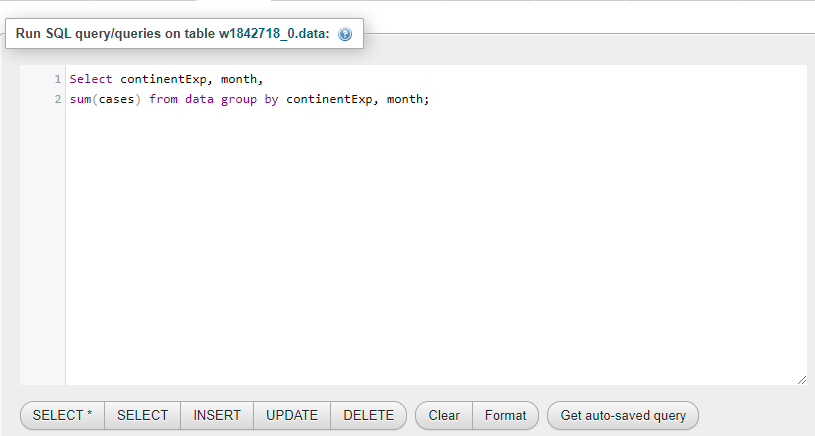
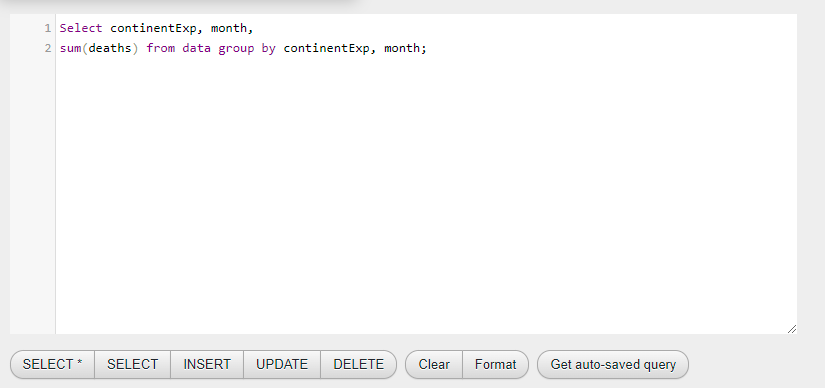
1. **Implementation of Materialized View**

A Materialized View perseveres the data returned from the view definition query and repeatedly gets updated as data variations in the fundamental tables.

**5.1 For each region and month report the COVID-19 cases and deaths, Intensive care unit cases as reported by different sources and vaccination rates for each vaccine per target group.**

Hereby have provided sample for the code and its output against the frequent. Remaining code is attached.

Code for ‘region and month’ with respect to Cases Code for ‘region and month’ with respect to Cases



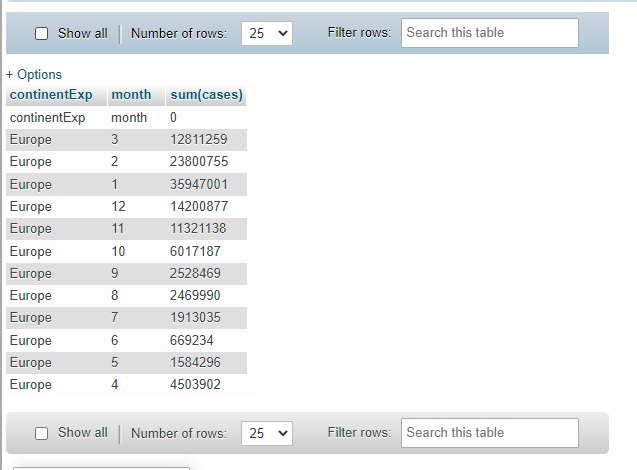
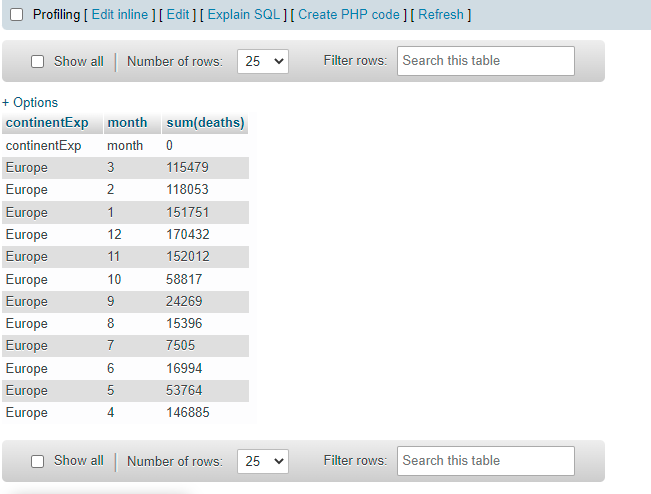


Table 2 Table 3

Entire code -

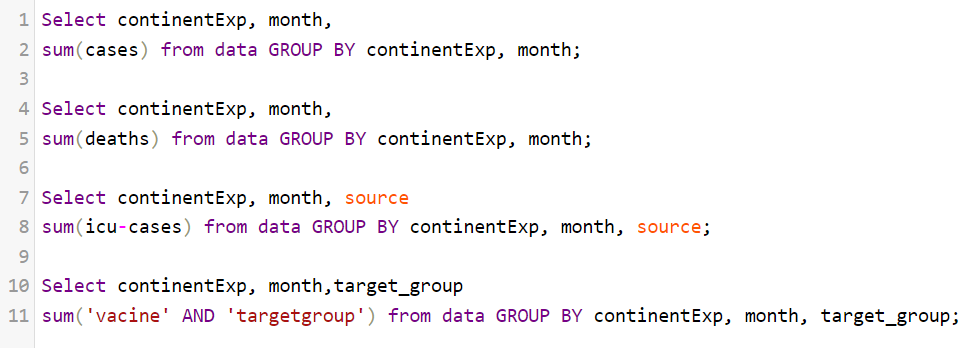


Table 4

**5.2 For each country and quarter report the COVID-19 cases and deaths, Intensive care unit cases and vaccination rates per target group.**

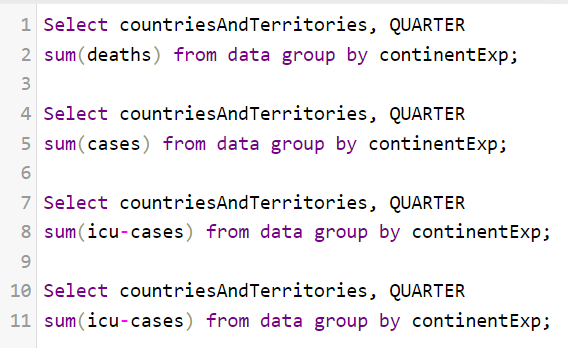
****

Table 5

**5.3 For each country and year report the COVID-19 cases and deaths, Intensive care unit cases and vaccination rates for the plus 60+ group.**

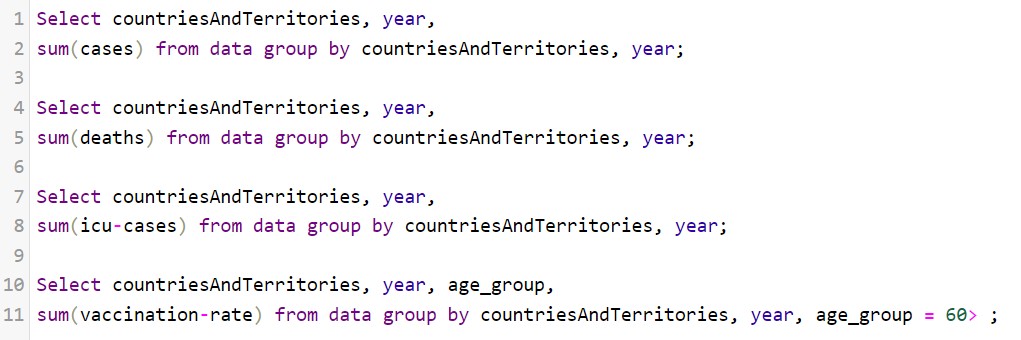


Table 6

The code for frequently asked Queries can be found in the attached file –

