Covid-19 Data Analysis using SQL

Section 1 - ER Model (20 points)

Draw an ER model for the following Business Requirement. Design an ER Model using the link https://erdplus.com/ or just use a pen and paper to design one!

You can **upload the images** of your ER Model and Diagram into the Airtable link here, you can **upload multiple files**, just follow the **naming convention**
https://airtable.com/appkB4tLGaUIf7rIf/shr4ggPuuyJqvEXYr

Make sure you only submit once! And follow the naming convention of the files to be accepted as a submission.

Business Requirement

When the COVID-19 virus began spreading uncontrollably in 2019, the world faced a pandemic. To survive and understand the impact of the pandemic, it became essential to analyze what the world went through.

To maintain a record of all cases worldwide, the world was divided into continents, and the total number of deaths for each continent was calculated. Each continent was assigned an ID, a name, and a record of total deaths to track this data effectively.

For further analysis, data was stored about each country. Since each continent consists of numerous countries, information such as total deaths, total cases, the date of the first recorded COVID case, and the date of the last recorded case was tracked for each country. Each country was also assigned a unique ID to provide a more detailed view of how COVID spread globally.

Additionally, each country was classified into an income class. Income classes were identified by a unique ID, the income class name (e.g., low, lower-middle, upper-middle, high), and the total deaths recorded within each class. This classification provided insights into the pandemic's impact on countries with varying income levels.

Finally, information about vaccines developed by each country was also recorded. Each vaccine was given a unique ID and name. Since multiple vaccines were developed by different countries, data was also collected on the number of individuals who received a specific vaccine (referred to as "survivors"). This allowed for tracking the number of survivors for each vaccine developed worldwide.

Section 2 - ER Diagram (10 points)

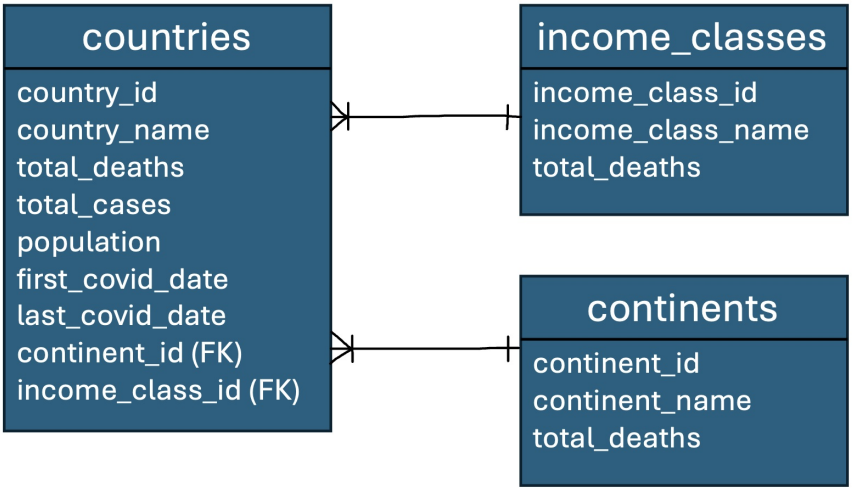
Convert the ER Model you made for the above requirement into an ER Diagram!

Make sure you **add the foreign keys** and **create new tables** if necessary.

Section 3 - Query the database!

COVID Database Structure

The database contains 3 tables, countries, continents and income_classes. Each table has it's own unique Id to identify each row and the following design.



info

Reset database

If you will be running any UPDATE queries, make sure you run this after running it.
Reset the database back to its original state before executing the code test below, click the "Reset covid Database" button below.

Examples

Type your answer in **sample.sql** file opened on your left and click on the **Check it!** button to check if your query generated the right output.

Example Question 1: Display all the information about all the countries.

Example Question 2: Find all the countries with more than 50000 deaths in total.

Questions

Question 1: Find all the countries with more than 100,000 covid cases.

Question 2: Find all countries where the first COVID case was reported in one month of COVID virus release. (COVID virus release date is ‘2019-12-31’).

Question 3: Find all the continents with starting with 'A'.

Question 4: Retrieve the first 5 countries based on their population in descending order.

Question 5: Find countries where total deaths is more than 10,000 and the population is less than 100 million.

Question 6: Find the top 3 countries with lowest number of total COVID cases and order them alphabetically.

Question 7: Group countries by continent_id and calculate the total population of each continent and name it as 'total_population'.

(2 columns only, first continent_id then total_population)

Question 8: Find the income class ids where the average total deaths as 'avg_deaths' in each income class is greater than 20000.

(2 columns only, income_class_id and avg_deaths).

Question 9: Calculate the total number of survivors in each country

(name the column as 'survivors') and list them in ascending order.

(2 columns only, country_name and survivors)

Question 10: Which country had the most survivors.

(2 columns only, country_name and survivors)

Question 11: Which country had the highest death rate.(rate is calculated by dividing total_deaths to population and multiplying it to 100)

(2 columns only, country_name and death_rate)

Question 12: Find all the countries with their names and the continent name they belong to.

(2 columns only, country_name and continent_name)

Question 13: Find the total number of countries in each income class.

(2 columns only, income_class_name and total_countries)

Question 14: Find the continent_name with the maximum total deaths.

(1 column, continent_name)

Question 15: Find which countries had more deaths than average deaths in the continent "Asia" (continent_id for Asia is 2).

(3 columns, country_name, total_deaths, continent_id)

Question 16: Find the income class with maximum deaths and what countries belong to that income class.

(3 columns, income_class_name, country_name, total_deaths)

Bonus Questions

If you lose marks in ER Modelling or ER Diagrams, then you can cover it up here!

Question 17: Building onto the previous question, find the population and death_rate of every country belonging to the income class with highest total deaths (We learnt to calculate death_rate in question 11).

(5 columns, income_class_name, country_name, total_deaths, population, death_rate)

Question 18: Just by looking at the output, which country has the highest death_rate?

Question 19: Do you really think economic wealth guarantee healthcare and immunity of the people? (Yes/No)

Note: Feel free to share your thoughts about the results you are seeing but keep in mind, this is just a small list of countries, containing majorly known countries of all the continents!