ISYS1055 (Practical) Database Concepts

Assessment 4: - Database Design Project

Part B: Designing the Database

The following information is given below from the given CSV data file from a global database of covid-19 vaccination:

Location:

Location, iso_code, vaccines, last_observation_date, source_name, source_website

Us_State_Vaccination:

Date, location, total_vaccinations, total_distributed, people_vaccinated, people_fully_vaccinated_per_hundred, total_vaccinations_per_hundred, people_vaccinated_per_hundred, distributed_per_hundred, daily_vaccinations_raw, daily_vaccinations, daily_vaccinations_per_million, share_doses_used, total_boosters, total_boosters_per_hundred.

Vaccination_By_age_group:

Location, date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred.

Vaccinations By Manufacturer:

Location, date, vaccine, total_vaccinations

Vaccinations:

Location, iso_code, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations_raw, daily_vaccinations, total_vaccinations_per_hundred, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, total_boosters_per_hundred, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred

Country_data/China:

Location, date, vaccine, source_url, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters

Country_data/India:

Location, date, vaccine, source_url, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters

Country_data/United_States:

Location, date, vaccine, source_url, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters

Country_data/ Ireland:

Location, date, vaccine, source_url, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters

Entity:

1. Country_Data
2. Vaccinations_Record
3. Vaccination By age group
4. Vaccination By manufacturer
5. US State Vaccinations
6. Vaccinations
7. Locations

Attributes names corresponding to their Entities are given below:

1. **Country_Data**: *(Country_Name/ID {pk}, location, date, vaccine, source_url, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters)*
2. **Locations**: *(Iso_code{pk}, location, vaccines, last_observation_date, source_name, source_website)*
3. **Vaccinations_by_age_groups**: *(age_group_id{pk}, location, date, age_group)*
4. **US_state_Vaccinations**: *(us_state_Id{pk}, date, location, total_vaccinations, total_distributed, people_vaccinated, people_fully_vaccinated_per_hundred, daily_vaccinations_raw, daily_vaccinations, share_doses_used, total_boosters, total_boosters_per_hundred)*
5. **Vaccinations_by_manufacturer**: *(manufacturer_ID {pk}, Location, date, vaccine, total_vaccinations)*
6. **Vaccines**: *(Vaccine_ID {pk}, location, name)*
7. **Vaccinations_Record**: *(Vaccinations_ID {pk}, location, iso_code, date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations_raw, daily_vaccinations)*

Relationships between the two entities are given below:

1. Country_Data and location: **One–to–Many**
2. Location and Vaccinations_by_age_groups: **One-to-Many**
3. Location and Vaccinations_by_manufacturer: **One-to-Many**
4. Location and Vaccinations_record: **One-to-Many**
5. Location and vaccines: **One-to-Many**
6. Location and data_source: **One-to-Many**
7. Location and US_State_Vaccinations: **One-to-Many**

8. Vaccinations_by_manufacturer and vaccinations record: **Many-to-Many**

Task 1: - Database ER diagram

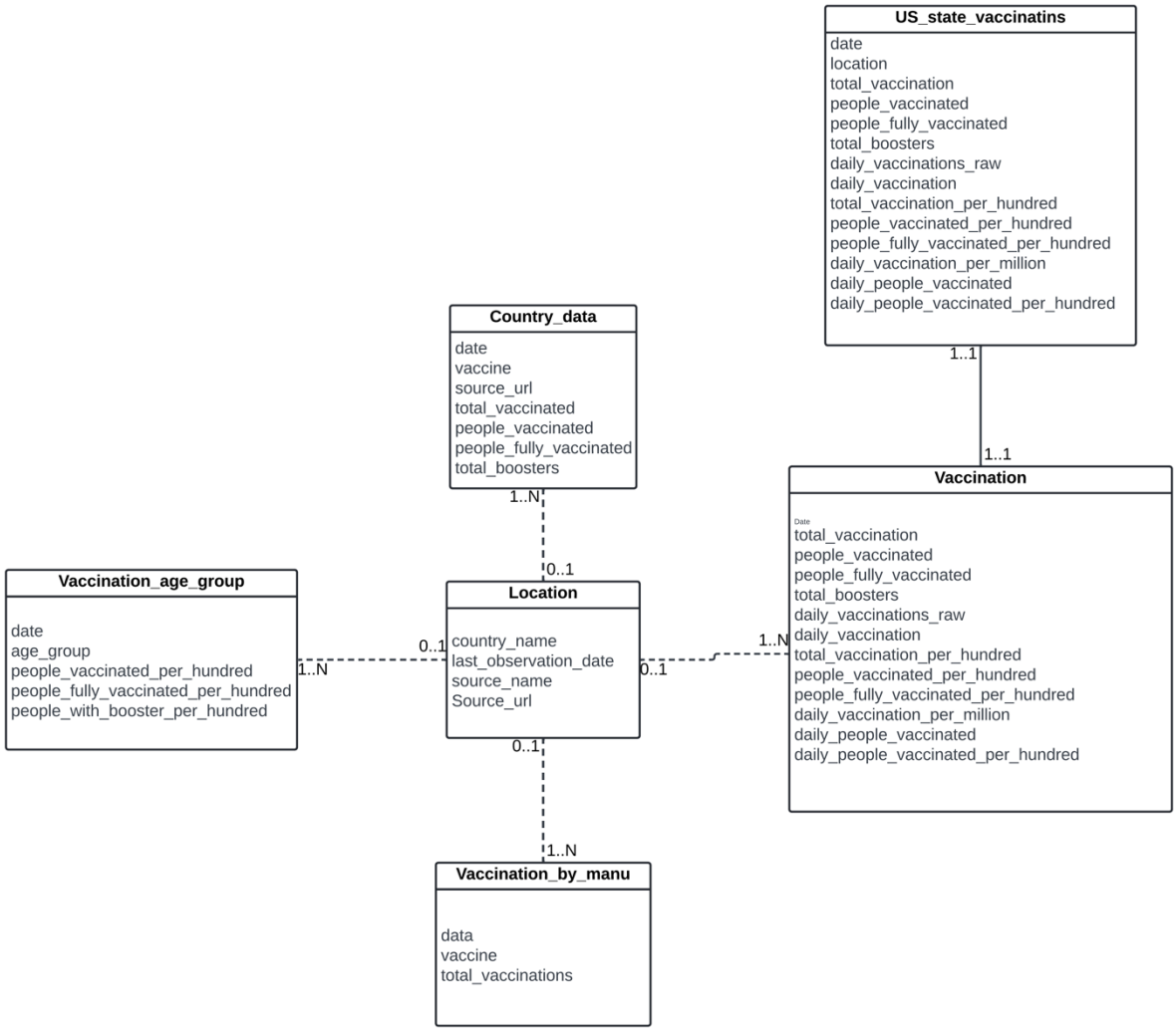

Fig 1: - Entity Relationship Diagram for Global Database of COVID-19 Vaccinations

Assumptions:

- Iso_code is the sole attribute that appears to have significant relationships with every other table if the location attribute is present in every table. With the exception of the US State Vaccinations chart, which lists locations at states rather than nations.
- The four nations—China, India, the United States, and Ireland—have all been merged into a single table in the ER diagram and given the name Country_Data in order to remove data redundancy and make the table simpler to read and understand.
    - A primary key called country_id has been added inside the County_Data table.
    - The foreign key has been identified as iso_code.
- It has been possible to view the various data sources that each country uses to determine the overall number of vaccines administered by creating a table called Data_Source.

    - The primary key data_source_id has been created.
    - The foreign key has been identified as iso_code.
- The ER diagram and database no longer have column names that finish in per_hundred or per_million, as those values were only the dividend values of the previous one. We have therefore eliminated such numbers to avoid data redundancy.

Task 2: - Explanation of Normalisation

By following the normalisation process, we have to develop the database schema:

Checking the table for 1NF: A relation or table is said to be in 1NF if that table shall not contain more than one value instead of an atomic value. This table or relation is not in 1NF since vaccination contains many attributes in location table. So, we further normalise into 1NF. We took the table location and created a new Vaccination table to normalize the problem to 1NF.

Checking the table for 2NF: It can be said that any table which is not violating the partial dependence rule must be in 2NF. A table or relation is said to be in 2NF when it has to be in 1NF, and the partial dependencies are not present.

Checking the table for 3NF: The table or relation must be in 2NF and there should not be any transitive dependency. As our table does not have any transitive dependencies, we can very well say that it is already in 3NF.

Task 3: - Database Schema

- Location: (Iso_code, location, vaccines, last_observation_date, source_name, source_website)
- Vaccinations: (us_state_id, vaccination*, date, location, total_vaccinations, total_distributed, people_vaccinated, people_fully_vaccinated, daily_vaccinations_raw, daily_vaccinations)
- Vaccinations_by_age_group: (age_group_id, location, iso_code*, date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred)
- Vaccinations_by_manufacturer: (manufacturer_id, location, iso_code*, date, vaccine, total_vaccinations)
- Country data: (country_id, location, iso_code*, date, vaccine, source_url, total_vaccinations, people_vaccinated, people_fully_vaccinated