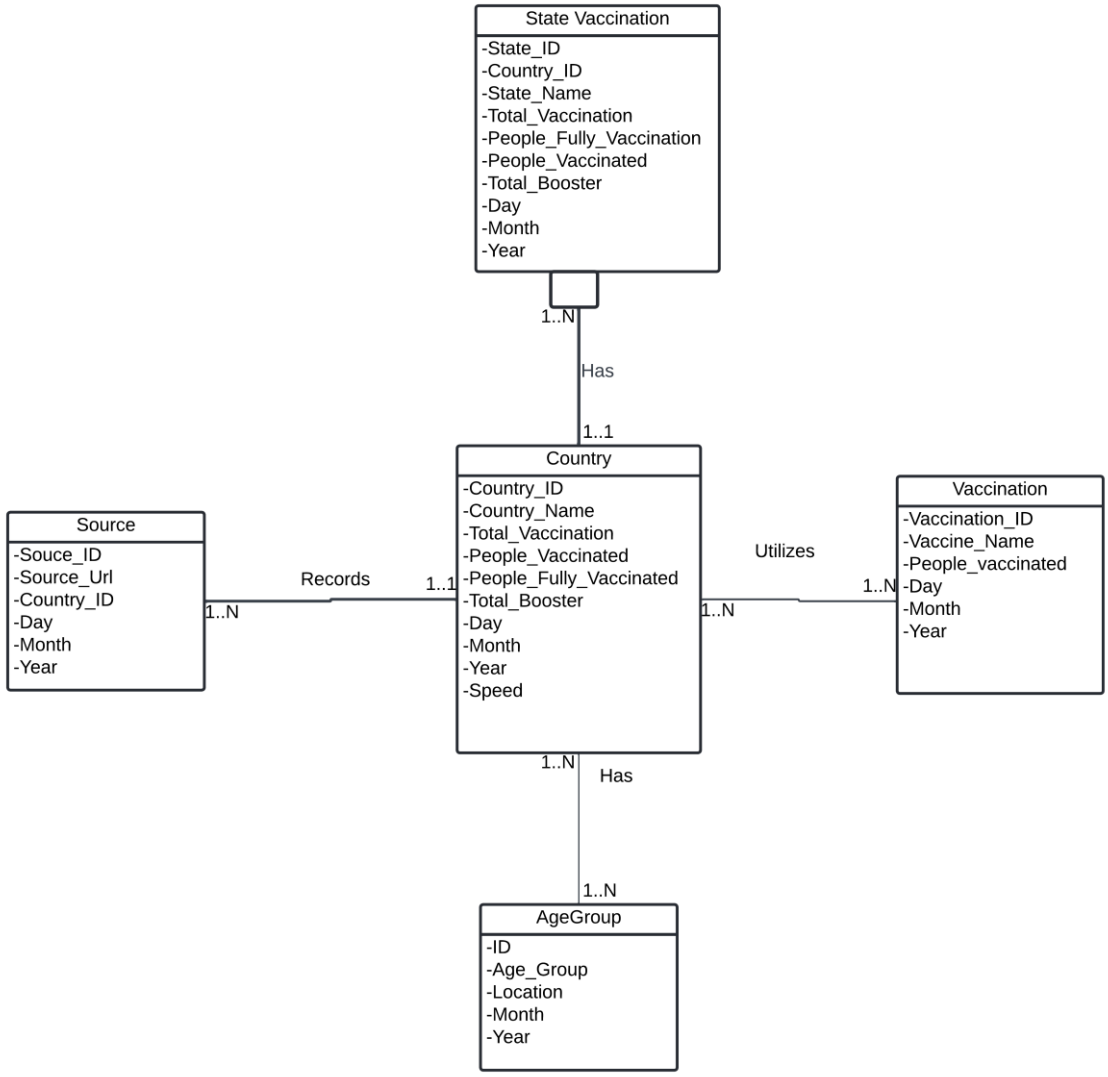
Part B: Designing the Database

Task B.1 Produce an ER diagram for a relational database that will be able to store the given dataset.

1. Database ER diagram



Assumptions:

- Each country's or state's vaccination data is unique.
- Vaccination details change over time, so the date is important.
- Speed refers to the difference in the total number of vaccinations between two consecutive days (yesterday and today) for a country. This measure helps us compare how quickly different countries are vaccinating their populations. The data for this calculation comes from the CSV files provided for four specific countries.

2. Explanation of Normalization Challenges and Resulting Changes

The challenges are:

- Data Redundancy: Potential for duplicate entries in country-specific vaccination data across multiple days if not normalized properly.
- Update Anomalies: Maintaining consistency across entries for countries and vaccinations could become cumbersome without appropriate normalization.

So, the resulting changes are:

- Source Entity: It has been isolated to eliminate redundancy and maintain a single reference point for each data source.
- To reduce the redundancy in date extraction, it has been divided to day, month, year in separate column.
- AgeGroup: It has been integrated through a relationship to the Country, rather than directly linking to Vaccination, to better structure demographic data handling.

3. Database Schema.

Step 1: Mapping strong entity

Country (Country_ID, Country_Name, Total_Vaccination, Total_Booster, Day, Month, Year)

Vaccination (Vaccination_ID, Vaccine_Name, People_Vaccinated, Day, Month, Year)

Source (Source_ID, Source_URl, Day, Month, Year, Country_ID*)

AgeGroup (ID, Age_Group, Location, Month, Year)

Step 2: Mapping the weak entity

State_Vaccination (State_ID, Country_ID*, State_Name, Total_Vaccination, People_Fully_Vaccinated, People_Vaccinated, Total_Booster, Day, Month, Year)

Step 3: Mapping 1:1 relationship

There no one to one relationship

Step 4: Mapping 1: N relationship

The relation of Country to State_Vaccination and Country to Source are one to many relationships and they are contains the primary key of country as foreign key so they are already mapped. No changes here in relationship since foreign keys have been properly assigned.

Step 5: Mapping many to many relationships

Utilizes (Country_ID*, Vaccination_ID*)

Has (Country_ID*, ID*)

Step 6: Mapping Multi Valued Attributes

Location of agegroup is multivalued so,

AgegroupLocation(ID*, Location)

Step 7: Mapping Higher degree relation

No, any higher degree relation.

The Final Schema is:

- AgegroupLocation(ID*, Location)
- Has (Country_ID*, ID*)
- Utilizes (Country_ID*, Vaccination_ID*)
- State_Vaccination (State_ID, Country_ID*, State_Name, Total_Vaccination, People_Fully_Vaccinated, People_Vaccinated, Total_Booster, Day, Month, Year)
- Country (Country_ID, Country_Name, Total_Vaccination, Total_Booster, Day, Month, Year)
- Vaccination (Vaccination_ID, Vaccine_Name, People_Vaccinated, Day, Month, Year)
- Source (Source_ID, Source_URl, Day, Month, Year, Country_ID*)

Normalization:

All the tables are in 1NF, 2NF and 3NF.