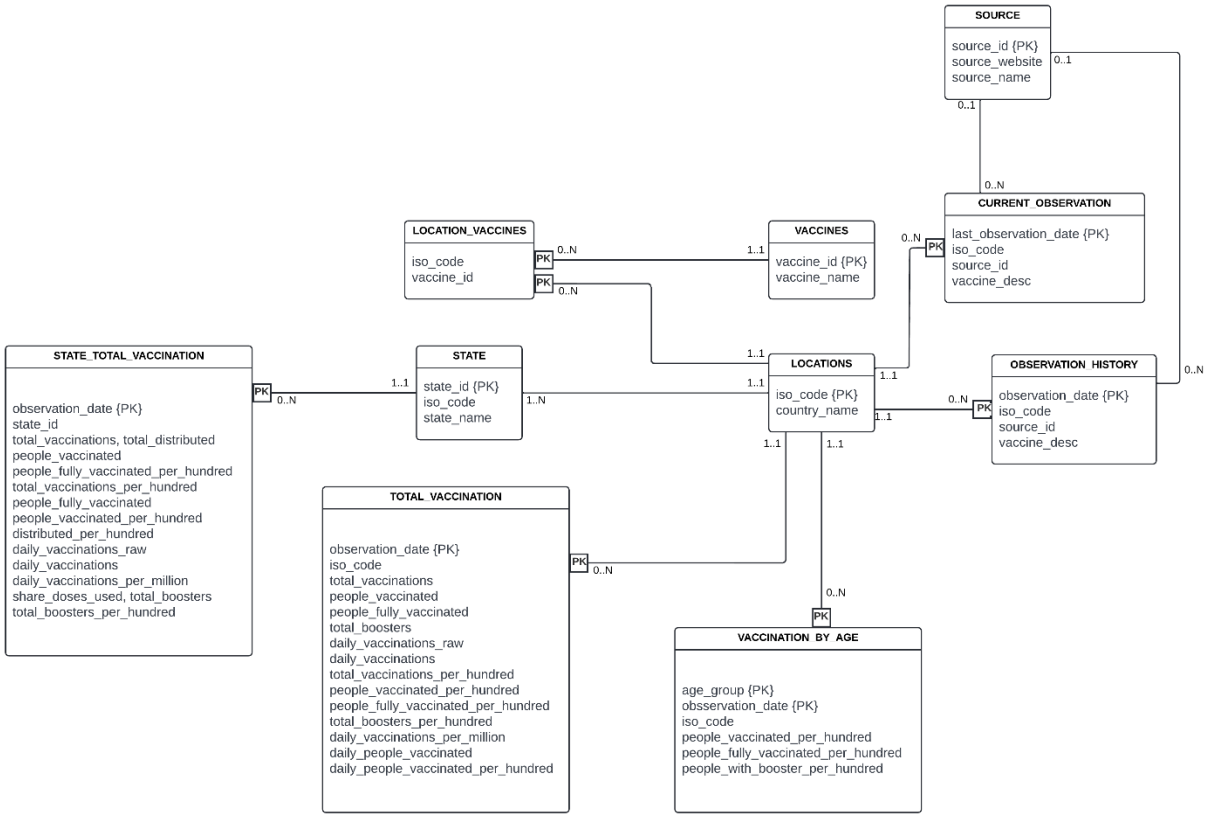
S3849118 | IVAN ALEGRIA

PART B: DESIGNING THE DATABASE

ER DIAGRAM:



Locations (location, iso_code, vaccines, last_observation_date, source_name, source_website)

Locations2 (location, iso_code, vaccines, last_observation_date, source_name, source_website)

Locations3 (iso_code, location)

Locations4 (vaccines, vaccine_name)

Locations5 (last_observation_date, iso_code*, source_name, source_website)

1NF as other non-primary key attributes are not functionally dependent on the composite key

Locations6 (last_observation_date, iso_code*, source_id*, vaccine_desc)

Locations7 (source_id, source_name, source_website)

Locations8 (iso_code*, vaccines*)

RESULTS =

LOCATIONS (iso_code, location)

3NF as there is only one non-primary key attribute

CURRENT_OBSERVATION (last_observation_date, iso_code*, source_id*, vaccine_desc)

Vaccine_desc displays the vaccine types administered in the observation date therefore is functionally dependent on the entire primary key

VACCINES (vaccine_id, vaccine_name)

3NF as there is only one non-primary key attribute

SOURCE (source_id, source_name, source_website)

3NF, assuming that there are various links in source_website for 1 source name.

LOCATION_VACCINES (iso_code*, vaccine_id*)

Vaccinations (location, iso_code, date, 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)

Vaccinations2 (location, iso_code*, date, 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)

Vaccinations3 (iso_code*, date, 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)

RESULTS =

TOTAL_VACCINATION (iso_code*, date, 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)

3NF as other non-primary key attributes cannot be calculated/have a transitive dependency on data not included in the database.

Vaccinations-by-manufacturer (location, date, vaccine, total_vaccinations)

Vaccinations-by-manufacturer2 (location, date, vaccine, total_vaccinations)

Vaccinations-by-manufacturer3 (vaccine_id, vaccine_name)

RESULTS =

VACCINES (vaccine_id, vaccine_name)

3NF as there is only one non-primary key attribute

Note Decided not include total_vaccinations as the data set were missing data for certain countries such as Australia, China, and England. And it would be difficult to manually input the total_vaccinations by manufacturer for the missing countries as the data specific for countries such as 'Australia' display the vaccines administered as a multi-valued attribute. Therefore, it would not be possibly to figure out how much of a specific vaccine was distributed without a dataset similar to the 'vaccinations-by-manufacturer'.

location	date	vaccine	source_url	total_vaccinations	people_vaccinated
Australia	2/21/2021	Pfizer/BioNTech	https://covidbaseau.com/	20	20

Vaccinations-by-age-group (location, date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

Vaccinations-by-age-group2 (location, date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

Vaccinations-by-age-group3 (iso_code*, date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

RESULTS =

VACCINATIONS_BY_AGE (iso_code*, date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

3NF as other non-primary key attributes cannot be calculated/have a transitive dependency on data not included in the database.

US_state_vaccinations (date location, total_vaccinations, total_distributed, people_vaccinated, people_fully_vaccinated_per_hundred, total_vaccinations_per_hundred, people_fully_vaccinated, 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)

US_state_vaccinations2 (date, location, total_vaccinations, total_distributed, people_vaccinated, people_fully_vaccinated_per_hundred, total_vaccinations_per_hundred, people_fully_vaccinated, 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)

US_state_vaccinations3 (date, total_vaccinations, total_distributed, people_vaccinated, people_fully_vaccinated_per_hundred, total_vaccinations_per_hundred, people_fully_vaccinated, 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)

US_state_vaccinations4 (State_id, iso_code*, state_name)

Created a new relation using the 'locations' data which had the information and names of all states located in the United States.

US_state_vaccinations3 (date, state_id*, total_vaccinations, total_distributed, people_vaccinated, people_fully_vaccinated_per_hundred, total_vaccinations_per_hundred, people_fully_vaccinated, 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)

RESULTS =

STATE_TOTAL_VACCINATIONS (observation_date, state_id*, total_vaccinations, total_distributed, people_vaccinated, people_fully_vaccinated_per_hundred, total_vaccinations_per_hundred, people_fully_vaccinated, 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)

STATE (State_id, iso_code*, state_name)

Australia, China, United Stated, and England Data:

Country (location, date, vaccine, source_url, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters)

Country2 (location, date, vaccine, source_url, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters)

Country3 (location, date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters)

Country4 (source_id, source_website, source_name)

Country5 (location, date, source_id*, vaccine_desc)

RESULTS =

Country3 (iso_code*, observation_date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters)

Table above uses the same primary keys as the table 'TOTAL_VACCINATIONS'. Therefore, will be merged to become one table

SOURCE (source_id, source_website, source_name)

OBSERVATION_HISTORY (iso_code*, observation_date, source_id*, vaccine_desc)

FINAL DATABASE SCHEMA:

LOCATIONS (iso_code, location)

CURRENT_OBSERVATION (last_observation_date, iso_code*, source_id*, vaccine_desc)

VACCINES (vaccines, vaccine_name)

SOURCE (source_id, source_name, source_website)

TOTAL_VACCINATION (iso_code*, observation_date, 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)

VACCINATIONS_BY_AGE (iso_code*, date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

STATE_TOTAL_VACCINATIONS (observation_date, state_id*, total_vaccinations, total_distributed, people_vaccinated, people_fully_vaccinated_per_hundred, total_vaccinations_per_hundred, people_fully_vaccinated, 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)

STATE (State_id, iso_code*, state_name)

OBSERVATION_HISTORY (iso_code*, observation_date, source_id*, vaccine_desc)

LOCATION_VACCINES (iso_code*, vaccine_id*)