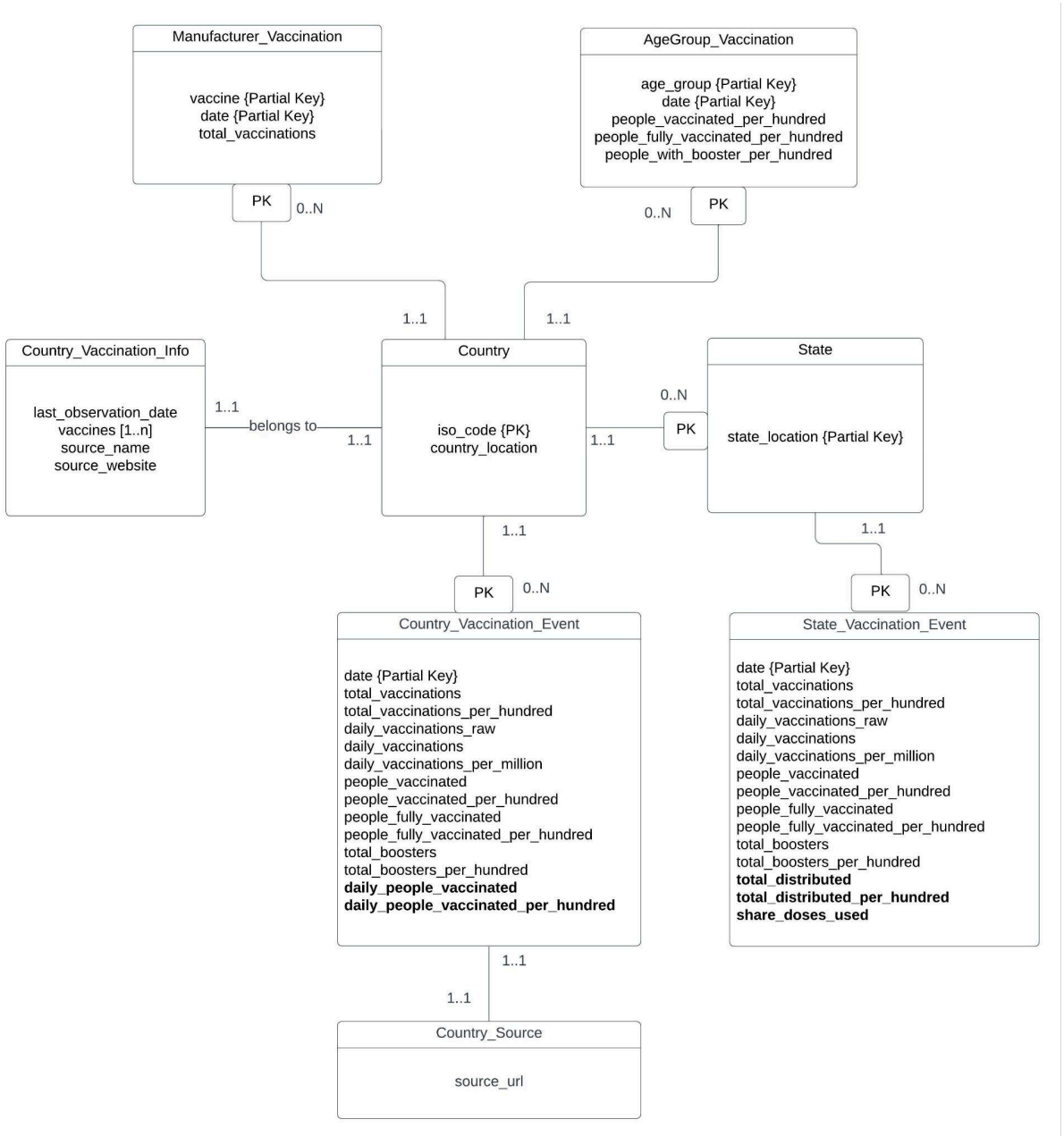
Assignment 4: Database Design Project

PART B

1. ER diagram



2. Database schema

Step 1: Map Strong Entities

Country (<u>iso_code,</u> country_location)

Step 2: Map Weak Entities

State (state_location, iso_code*)
Manufacturer_Vaccination (iso_code*, date, vaccine, total_vaccinations)
AgeGroup_Vaccination (iso_code*, date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

Country_Vaccination_Event(iso_code*, date, total_vaccinations, total_vaccinations_per_hundred, daily_vaccinations_raw, daily_vaccinations, daily_vaccinations_per_million, people_vaccinated, people_vaccinated_per_hundred, people_fully_vaccinated, people_fully_vaccinated_per_hundred, total_boosters, total_boosters_per_hundred, daily_people_vaccinated, daily_people_vaccinated_per_hundred)

Country_Source(iso_code*, date*, source_url)

State_Vaccination_Event(iso_code*, state_location*, date, total_vaccinations, total_vaccinations_per_hundred, daily_vaccinations_raw, daily_vaccinations, daily_vaccinations_per_million, people_vaccinated, people_vaccinated_per_hundred, people_fully_vaccinated, people_fully_vaccinated_per_hundred, total_boosters, total_boosters_per_hundred, total_distributed, total_distributed_per_hundred, share_doses_used)

Step 3: Map 1:1 Relationships

Country (iso_code, last_observation_date, source_name, source_website)

Step 4: Map 1:N Relationships

N/A

Step 5: Map M:N Relationships

N/A

Step 6: Multi-valued Attributes

Country_Vaccine(iso_code*, vaccine)

Step 7: Map Higher-Degree Relationships

N/A

Final Schema

Country (iso_code, last_observation_date, source_name, source_website)

State (state_location, iso_code*)

Manufacturer_Vaccination (iso_code*, date, vaccine, total_vaccinations)

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

Country_Vaccination_Event(iso_code*, date, total_vaccinations, total_vaccinations_per_hundred, daily_vaccinations_raw, daily_vaccinations, daily_vaccinations_per_million, people_vaccinated, people_vaccinated_per_hundred, people_fully_vaccinated, people_fully_vaccinated_per_hundred, total_boosters, total_boosters_per_hundred, daily_people_vaccinated, daily_people_vaccinated_per_hundred)

Country_Source(iso_code*, date*, source_url)

Country_Vaccine(iso_code*, vaccine)

State_Vaccination_Event(iso_code*, state_location*, date, total_vaccinations, total_vaccinations_per_hundred, daily_vaccinations_raw, daily_vaccinations, daily_vaccinations_per_million, people_vaccinated, people_vaccinated_per_hundred, people_fully_vaccinated, people_fully_vaccinated_per_hundred, total_boosters, total_boosters_per_hundred, total_distributed, total_distributed_per_hundred, share_doses_used)

3. 3NF normalisation

Step 1: Check for 1NF

To be in 1NF, all attributes must contain atomic values. From the given schema, all attributes appear atomic, so the schema fulfills 1NF.

Step 2: Check for 2NF

To be in 2NF, the schema must be in 1NF and all non-key attributes must be fully dependent on the entire primary key. The dataset fulfills this requirement.

Step 3: Check for 3NF

Transitive Dependencies: There are attributes in State_Vaccination_Event and Country_Vaccination_Event that might depend on other non-primary key attributes through calculations rather than the composite keys themselves.

1. State_Vaccination_Event

- **total_vaccinations_per_hundred** and **people_vaccinated_per_hundred** are dependent on total_vaccinations and people_vaccinated, respectively. These attributes are derived and calculated based on the total population of the state.
- **daily_vaccinations_per_million** is based on daily_vaccinations in relation to the state's population.

- **total_distributed_per_hundred** is calculated based on the total_distributed in relation to the state's population.
- **share_doses_used** depends on both total_vaccinations and total_distributed.

2. Country_Vaccination_Event

- **total_vaccinations_per_hundred** depends on total_vaccinations and the total population of the country.
- **daily_vaccinations_per_million** depends on daily_vaccinations and the total population of the country.
- **people_vaccinated_per_hundred** depends on people_vaccinated and the total population of the country.
- **people_fully_vaccinated_per_hundred** depends on people_fully_vaccinated and the total population of the country.
- **total_boosters_per_hundred** depends on total_boosters and the total population of the country.
- **daily_people_vaccinated_per_hundred** depends on daily_people_vaccinated and the total population of the country.

Decompose into 3NF normalization

Separate the raw metrics of State_Vaccination_Event

State_Vaccination_Metrics (iso_code*, state_location*, date, total_vaccinations, daily_vaccinations_raw, daily_vaccinations, people_vaccinated, people_fully_vaccinated, total_distributed, total_boosters)

Separate the calculated metrics of State_Vaccination_Event

State_Calculated_Metrics (iso_code*, state_location*, date, total_vaccinations_per_hundred, daily_vaccinations_per_million, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, total_distributed_per_hundred, share_doses_used, total_boosters_per_hundred)

Separate the raw metrics of Country_Vaccination_Event

Country_Vaccination_Metrics (iso_code*, date, total_vaccinations, daily_vaccinations_raw, daily_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_people_vaccinated)

Separate the calculated metrics of Country_Vaccination_Event

Country_Calculated_Metrics (iso_code*, date, total_vaccinations_per_hundred, daily_vaccinations_per_million, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, total_boosters_per_hundred, daily_people_vaccinated_per_hundred)

After the 3NF normalization, the dataset gains:
Preserved Data Integrity: By keeping non-key attributes dependent only on the primary key, aligning with 3NF.
Enhanced Maintainability and Flexibility: By supporting recalculation of calculated metrics without altering raw metrics.

Schema

Country(iso_code, country_location, last_observation_date, source_name, source_website)

Country_Source(iso_code*, date*, source_url)

Country_Vaccine(iso_code*, vaccine)

State (state_location, iso_code*)

Manufacturer_Vaccination (iso_code*, date, vaccine, total_vaccinations)

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

State_Vaccination_Metrics (iso_code*, state_location*, date, total_vaccinations, daily_vaccinations_raw, daily_vaccinations, people_vaccinated, people_fully_vaccinated, total_distributed, total_boosters)

State_Calculated_Metrics (iso_code*, state_location*, date, total_vaccinations_per_hundred, daily_vaccinations_per_million, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, total_distributed_per_hundred, share_doses_used, total_boosters_per_hundred)

Country_Vaccination_Metrics (iso_code*, date, total_vaccinations, daily_vaccinations_raw, daily_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_people_vaccinated)

Country_Calculated_Metrics (iso_code*, date, total_vaccinations_per_hundred, daily_vaccinations_per_million, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, total_boosters_per_hundred, daily_people_vaccinated_per_hundred)