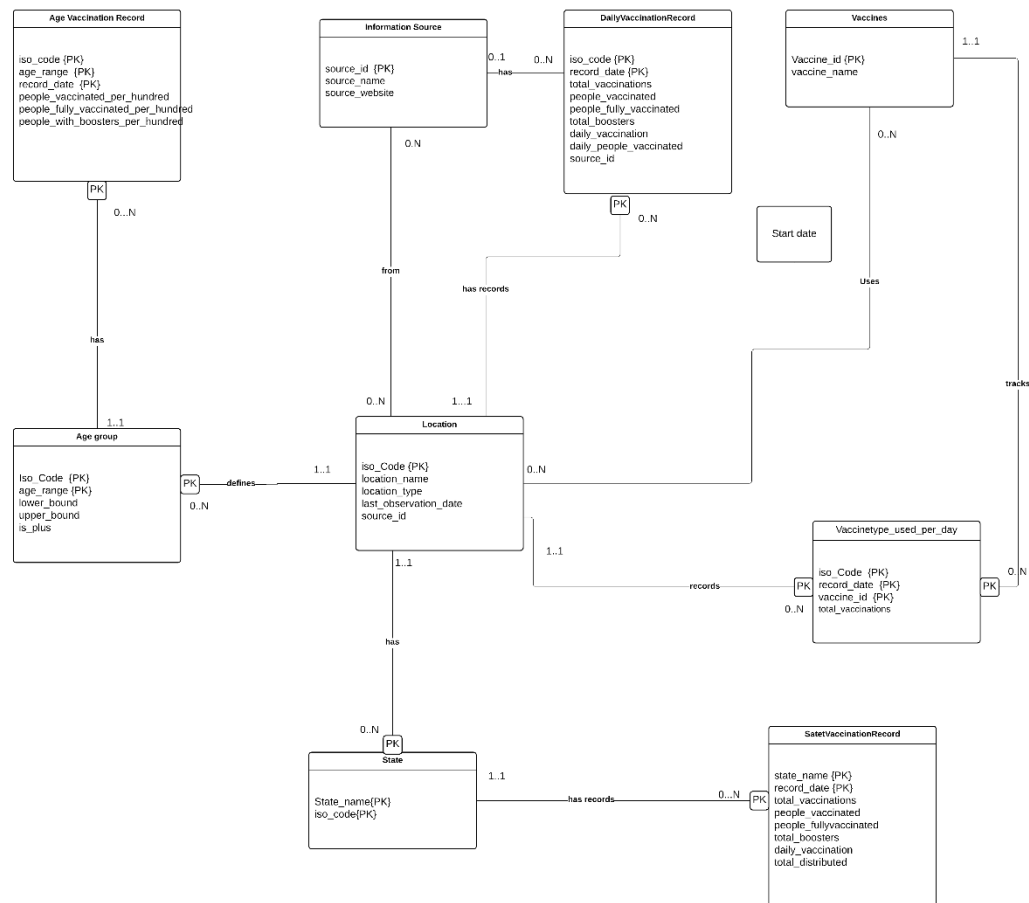


DATABASE CONCEPTS ASSIGNMENT 4

Part B

1) Designing an Entity-Relationship Diagram



Assumptions Made

1) Information source

- 1) Each information source may be associated with at least 0 or at most many daily vaccinations record
- 2) Each information source can be associated with at least 0 or at most many locations

2) Location

- 1) Each location may be associated with at least 0 or at most many information source

- 2) Each location may be associated with at least 0 or at most many daily vaccinations record
- 3) Each location may be associated with at least 0 or at most many vaccines types
- 4) Each location may be associated with at least 0 or at most many vaccines types used per day record
- 5) Each location may be associated with at least 0 or at most many age groups
- 6) Each location may be associated with at least 0 or at most many states

3)State

- 1) Each state may be associated with at least 1 and at most 1 country
- 2) 1) Each state may be associated with at least 0 or at most many daily state vaccination records

4)Daily Vaccination Record

- 1) Each vaccination activity may be associated with at 1 location
- 2) Each vaccination activity may be associated with at least 0 or at most 1 information source

5)Vaccine

- 1) Each vaccine may be associated with at least 0 or at most many locations
- 2) Each vaccine may be associated with at least 0 or at most many vaccination type used per day

6)Age Group

- 1) Each age group can be associated with at least and at most 1 location
- 2) Each age group can be associated with at least 0 and at most many age vaccination record

7)Age vaccination record

- 1) Each record can be associated with at least and at most 1 age group

8)State Vaccination Record

- 1) Each vaccination activity may be associated with at 1 location

9)Vaccination used per day (per country)

- 1) Each daily vaccination total may be associated with at 1 state
- 2) Each vaccination activity may be associated with at 1 vaccine type

2)Normalization Challenges and Resulting Changes

The original schema aimed to capture vaccination data by country, but as the requirements expanded, it needed to handle data at multiple geographic levels (e.g., region, country, global). This led to several challenges, prompting adjustments to meet 3NF requirements and ensure efficient, flexible data management.

Challenge 1: Geographic Levels in Location

Initially, I used a Country table, but later expanded it to a more flexible Location table to store regional, national, and global data. To differentiate these, I added a location_type attribute, making it easier to filter records by geographic level and avoid redundant data.

Challenge 2: Avoiding Redundancy Across Tables

I initially planned to store location and source information directly in tables like Daily Vaccination Record and State Vaccination Record. To avoid redundancy, I introduced Location and Information Source tables, linking them through iso_code and source_id. This ensured consistency and central management of these attributes.

Challenge 3: Tracking Vaccine Types

To manage specific vaccines administered by location and date, I separated this data into Vaccine Type Used Per Day Per Country instead of embedding it in the main vaccination record. This kept the schema normalized and simplified querying vaccine-specific data over time.

Challenge 4: Age Group Data Management

Rather than storing age-specific data directly in Daily Vaccination Record, I created an Age_group table to manage age ranges and keep the schema flexible. This change eliminated partial dependencies and improved adaptability.

Challenge 5: Standardizing Sources

Initially, source information was embedded across tables, risking inconsistency. Moving it to Information Source and linking it via source_id through a many-to-many Location Info Source table provided a centralized, consistent structure.

Summary of Changes:

1. Location Type: Replaced Country with Location and added location_type for various geographic levels.
2. Normalization with Reference Tables: Introduced Location Info Source and Vaccine Type Used Per Day Per Country for reduced redundancy.
3. Age Group Table: Created a separate Age_group table for age-specific vaccination data.
4. Centralized Source Management: Standardized sources through Information Source, ensuring consistency across the schema.

These changes resulted in a fully normalized, flexible schema that can efficiently manage complex vaccination data requirements across multiple regions.

3) Schema

Step 1

Strong Entities

Location (iso_code, location_name, location_type, last_observation_date, source_id)

Vaccine (vaccine_id, vaccine_name)

Information Source (source_id, source_name, source_website)

Step 2

Weak Entity

State (state_name, iso_code*)

Age_group (iso_code*, age_range, lower_bound, upper_bound, is_plus)

Daily Vaccination Record (iso_code*, record_date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations, source_id)

State Vaccination record (state_name*, record_date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations, total_distributed)

Age Vaccination Record(age_range*, iso_code*, record_date, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_boosters_per_hundred)

Vaccine type used per day per country (iso_code*, record_date, vaccine_id*, total_vaccinations)

Step 3

Map 1:1 relationship

N/A

Step 4

Map 1:M relationship

State (state_name, iso_code*)

Age_group (iso_code*, age_range, lower_bound, upper_bound, is_plus)

Daily Vaccination Record (iso_code*, record_date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations, source_id)

Vaccine type used per day per country (iso_code*, record_date, vaccine_id*, total_vaccinations)

State Vaccination record (state_name*, record_date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations, total_distributed)

Age Vaccination Record (age_range*, iso_code*, record_date, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_boosters_per_hundred)

Step 5

Map M:M relationship

Location wise vaccine (iso_code*, vaccine_id*, start_date)

Location Info source (iso_code*, source_id*)

Step 6

Map Multi valued attributes

No further breakdown required

Step 7

Higher order relationship

No further breakdown required

Step 8

Recursive relation

No action required

Final Schema

Location (iso_code, location_name, location_type, last_observation_date, source_id)

Vaccine (vaccine_id, vaccine_name)

Information Source (source_id, source_name, source_website)

State (state_name, iso_code*)

Age_group (iso_code*, age_range, lower_bound, upper_bound, is_plus)

Daily Vaccination Record (iso_code*, record_date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations, source_id)

Vaccine type used per day per country (iso_code*, record_date, vaccine_id*, total_vaccinations)

State Vaccination record (state_name*, record_date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations, total_distributed)

Age Vaccination Record (age_range*, iso_code*, record_date, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_boosters_per_hundred)

Location wise vaccine (iso_code*, vaccine_id*, start_date)

Location Info source (iso_code*, source_id*)

Normalisation Check

Step 1: Identify Functional Dependencies (FDs)

1. Location:
 - iso_code → location_name, location_type, last_observation_date, source_id
2. Vaccine:
 - vaccine_id → vaccine_name
3. Information Source:
 - source_id → source_name, source_website
4. State:
 - iso_code, state_name → unique (assuming iso_code can belong to multiple states but each state-iso_code pair is unique)
5. Age_group:
 - iso_code, age_range → lower_bound, upper_bound, is_plus
6. Daily Vaccination Record:
 - iso_code, record_date → total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations, source_id
7. Vaccine Type Used Per Day Per Country:
 - iso_code, record_date, vaccine_id → total_vaccinations
8. State Vaccination Record:
 - state_name, record_date → total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations, total_distributed
9. Age Vaccination Record:
 - age_range, iso_code, record_date → people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_boosters_per_hundred
10. Location Wise Vaccine:
 - iso_code, vaccine_id → start_date
11. Location Info Source:
 - iso_code, source_id (assumed as unique combinations per location and source)

Step 2: Check for 1NF (First Normal Form)

1NF requires that all values in a table are atomic and each attribute has a single value.

- All tables seem to have atomic values with no repeating groups.
- Hence, the schema is in 1NF.

Step 3: Check for 2NF (Second Normal Form)

2NF requires the schema to be in 1NF, and every non-key attribute must depend on the whole primary key (no partial dependency).

1. For tables with a single-attribute primary key (e.g., Vaccine, Information Source), there are no partial dependencies.
2. For tables with composite primary keys (e.g., Daily Vaccination Record), each non-key attribute depends on the entire composite key.

Since no partial dependencies exist, the schema is in 2NF.

Step 4: Check for 3NF (Third Normal Form)

3NF requires the schema to be in 2NF, and every non-key attribute must depend only on the primary key (no transitive dependencies).

1. Location:
 - iso_code is the primary key, and all other attributes (location_name, location_type, last_observation_date, source_id) depend directly on it. 3NF satisfied.
2. Vaccine:
 - vaccine_id is the primary key, and vaccine_name depends directly on it. 3NF satisfied.
3. Information Source:
 - source_id is the primary key, and source_name, source_website depend directly on it. 3NF satisfied.
4. State:
 - Composite primary key: (state_name, iso_code). All other attributes depend on this composite key directly, and no transitive dependencies exist. 3NF satisfied.
5. Age_group:
 - Composite primary key: (iso_code, age_range). lower_bound, upper_bound, and is_plus depend directly on this key. 3NF satisfied.
6. Daily Vaccination Record:
 - Composite primary key: (iso_code, record_date). All other attributes (total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations, source_id) depend directly on this key. 3NF satisfied.
7. Vaccine Type Used Per Day Per Country:

- Composite primary key: (iso_code, record_date, vaccine_id). total_vaccinations depends directly on this key. 3NF satisfied.
8. State Vaccination Record:
- Composite primary key: (state_name, record_date). All other attributes depend directly on this key. 3NF satisfied.
9. Age Vaccination Record:
- Composite primary key: (age_range, iso_code, record_date). All other attributes depend directly on this key. 3NF satisfied.
10. Location Wise Vaccine:
- Composite primary key: (iso_code, vaccine_id). start_date depends directly on this key. 3NF satisfied.
11. Location Info Source:
- Composite primary key: (iso_code, source_id). No transitive dependencies, as the key is simple. 3NF satisfied.

Conclusion

Since each table is in 3NF, with no partial or transitive dependencies and all non-key attributes depending solely on the primary keys, schema is in 3NF.