

# **ISYS1055/1057/3412 (Practical) Database Concepts**

## **Assessment 4: Database Design Project**

**Course Name:** Database Concepts (2250)

**Student Name:** Antony Rosario John Peter

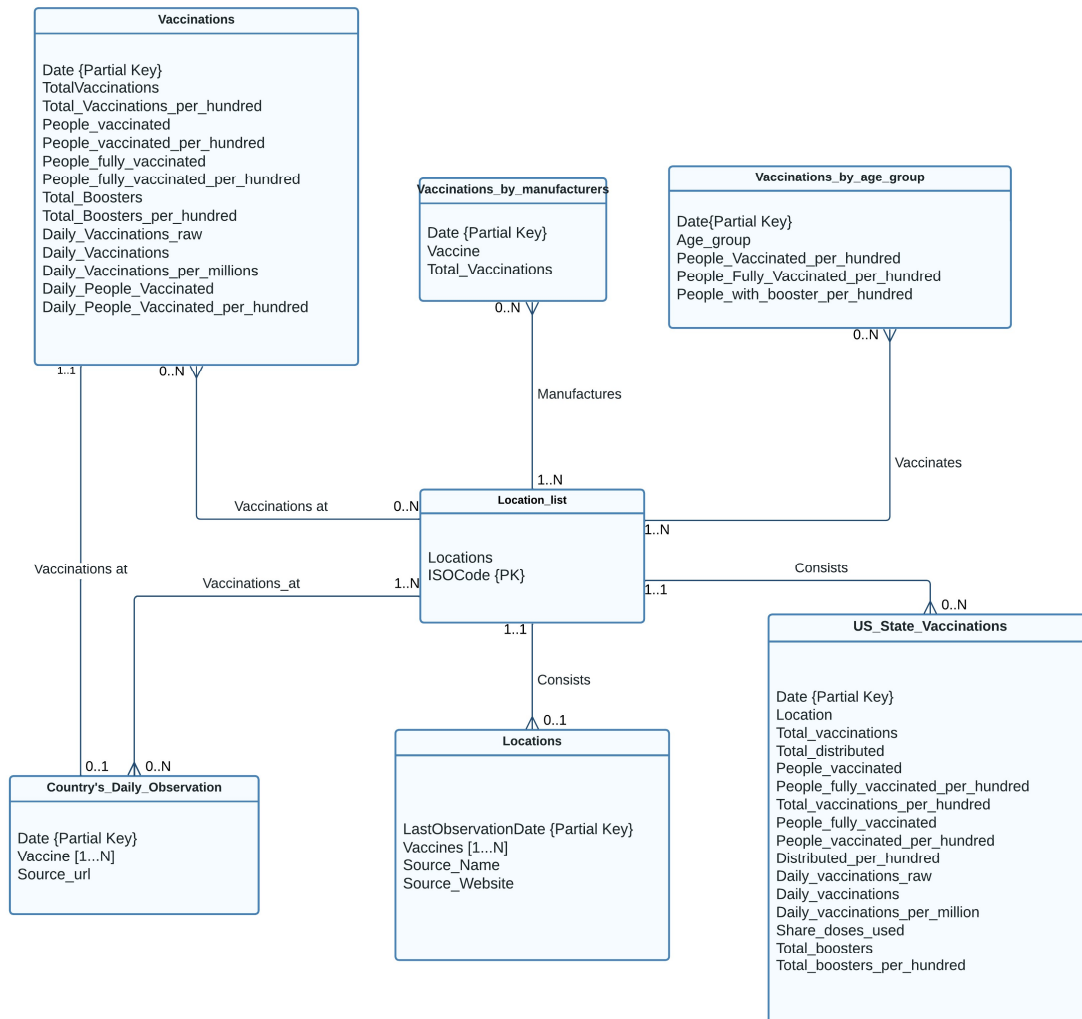
**Student ID:** S3940203

**Student Email:** s3940203@student.rmit.edu.au

## Part B: Designing the Database (10%)

### Task B.1: Designing an Entity Relationship Model

Entity Relationship Model for Data on COVID-19 (coronavirus) vaccinations by Our World



#### Assumptions:

- As given in the requirement, we need to have an entity of four countries, Australia, Germany, Italy, and United States, but we have the same details in the Vaccinations Entity
- I have created an entity called Country's\_Daily\_observation and have combined the data of Australia, Germany, Italy and United States into single table and rest of the information can be fetched from Vaccinations Entity
- I have created a new entity called Location\_list with Locations and ISO\_Code as attributes where ISO\_Code will be the primary key of the entity

- Instead of having the location names in the entities, I have planned to replace them with ISO\_Code
- Here, the count of location name in Locations entity and the count of unique location names in Vaccinations does not equal, So I have taken the ISO\_Code from Vaccinations entity as well to create the Location\_list entity
- We have multi-valued attributes in some entities will be managed in later part of the Schema

## **Task B.2: Mapping an ER Model to a Relational Database Schema**

### Step 1: Strong Entities

- Location\_list (ISO\_Code, Locations)

### Step 2: Weak Entities

- Locations (ISO\_Code\*, LastObservationDate, Vaccines [1...N], Source\_Name, Source\_Website)
- Vaccinations (ISO\_Code\*, Date, TotalVaccinations, Total\_Vaccinations\_per\_hundred, People\_vaccinated, People\_vaccinated\_per\_hundred, People\_fully\_vaccinated, People\_fully\_vaccinated\_per\_hundred, Total\_Boosters, Total\_Boosters\_per\_hundred, Daily\_Vaccinations\_raw, Daily\_Vaccinations, Daily\_Vaccinations\_per\_millions, Daily\_People\_Vaccinated, Daily\_People\_Vaccinated\_per\_hundred)
- Vaccinations\_by\_manufacturers (ISO\_Code\*, Date, Vaccines, Total\_Vaccinations)
- Vaccinations\_by\_age\_group (ISO\_Code\*, Date, Age\_group, People\_Vaccinated\_per\_hundred, People\_Fully\_Vaccinated\_per\_hundred, People\_with\_booster\_per\_hundred)
- US\_State\_Vaccinations (ISO\_Code\*, 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)
- Country's\_Daily\_Observation (ISO\_Code\*, Date, Vaccine [1...N], Source\_url)

### Step 3: One-to-One Relationships

Locations has a 1...1 relationship with Location\_list, but ISO\_Code is already included in Locations as a foreign key, together with LastObservationDate makes the primary key of Locations.

- Locations (ISO\_Code\*, LastObservationDate, Vaccines [1...N], Source\_Name, Source\_Website)

US\_State\_Vaccinations has a 1...1 relationship with Location\_list, but ISO\_Code is already included in US\_State\_Vaccinations as a foreign key, together with Date, Location makes the primary key of US\_State\_Vaccinations.

- US\_State\_Vaccinations (ISO\_Code\*, 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)

### Step 4: One-to-Many Relationships

As we have already included ISO\_Code as foreign key with 1...N relationship, so no further action is required.

### Step 5: Many-to-Many Relationships

Nothing to do here, we don't have any Many to Many relationship

### Step 6: Multi-valued Attributes

As we can see in Locations, Vaccinations\_by\_age\_group and Country's\_Daily\_Observation have Multi-valued Attributes of Vaccines

So, I have created a new entity named Vaccine\_list, which consist of all the unique types of vaccines which are given in above listed four entities and have created a Vaccine\_ID for all the individual Vaccines.

- Vaccine\_list (Vaccine\_ID, Vaccine\_Name)

### Step 7: Higher Degree Relationships

Nothing to do here, we do not have any ternary relationship

### Special Case:

Nothing to do here, we do not have any Special Cases.

### Relational Database Schema:

- Location\_list (ISO\_Code, Locations)
- Locations (ISO\_Code\*, LastObservationDate, Vaccines [1...N], Source\_Name, Source\_Website)
- Vaccinations (ISO\_Code\*, Date, TotalVaccinations, Total\_Vaccinations\_per\_hundred, People\_vaccinated, People\_vaccinated\_per\_hundred, People\_fully\_vaccinated, People\_fully\_vaccinated\_per\_hundred, Total\_Boosters, Total\_Boosters\_per\_hundred, Daily\_Vaccinations\_raw, Daily\_Vaccinations, Daily\_Vaccinations\_per\_millions, Daily\_People\_Vaccinated, Daily\_People\_Vaccinated\_per\_hundred)
- Vaccinations\_by\_manufacturers (ISO\_Code\*, Date, Vaccine, Total\_Vaccinations)
- Vaccinations\_by\_age\_group (ISO\_Code\*, Date, Age\_group, People\_Vaccinated\_per\_hundred, People\_Fully\_Vaccinated\_per\_hundred, People\_with\_booster\_per\_hundred)
- US\_State\_Vaccinations (ISO\_Code\*, 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)
- Country's\_Daily\_Observation (ISO\_Code\*, Date, Vaccine [1...N], Source\_url)
- Vaccine\_list (Vaccine\_ID, Vaccine\_Name)

### Normalisation:

#### 1. Functional Dependencies:

- ISO\_Code -> Locations
- ISO\_Code, LastObservationDate -> Vaccines [1...N], Source\_Name, Source\_Website
- ISO\_Code, Date -> TotalVaccinations, Total\_Vaccinations\_per\_hundred, People\_vaccinated, People\_vaccinated\_per\_hundred, People\_fully\_vaccinated, People\_fully\_vaccinated\_per\_hundred, Total\_Boosters, Total\_Boosters\_per\_hundred, Daily\_Vaccinations\_raw, Daily\_Vaccinations, Daily\_Vaccinations\_per\_millions, Daily\_People\_Vaccinated, Daily\_People\_Vaccinated\_per\_hundred
- ISO\_Code, Date, Vaccines -> Total\_Vaccinations

- ISO\_Code, Date, Age\_group -> People\_Vaccinated\_per\_hundred, People\_Fully\_Vaccinated\_per\_hundred, People\_with\_booster\_per\_hundred
- ISO\_Code, 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
- ISO\_Code, Date -> Vaccine [1...N], Source\_url
- Vaccine\_ID -> Vaccine\_Name

## 2. Highest Normal Form:

- Location\_list (ISO\_Code, Locations) -> 3NF
- Locations (ISO\_Code\*, LastObservationDate, Vaccines [1...N], Source\_Name, Source\_Website)

-- A Non primary key attribute has a multi-valued attribute, failed 1NF

- Vaccinations (ISO\_Code\*, Date, TotalVaccinations, Total\_Vaccinations\_per\_hundred, People\_vaccinated, People\_vaccinated\_per\_hundred, People\_fully\_vaccinated, People\_fully\_vaccinated\_per\_hundred, Total\_Boosters, Total\_Boosters\_per\_hundred, Daily\_Vaccinations\_raw, Daily\_Vaccinations, Daily\_Vaccinations\_per\_millions, Daily\_People\_Vaccinated, Daily\_People\_Vaccinated\_per\_hundred) -> 3NF
- Vaccinations\_by\_manufacturers (ISO\_Code\*, Date, Vaccine, Total\_Vaccinations) -> 3NF
- Vaccinations\_by\_age\_group (ISO\_Code\*, Date, Age\_group, People\_Vaccinated\_per\_hundred, People\_Fully\_Vaccinated\_per\_hundred, People\_with\_booster\_per\_hundred) -> 3NF
- US\_State\_Vaccinations (ISO\_Code\*, 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) -> 3NF
- Country's\_Daily\_Observation (ISO\_Code\*, Date, Vaccine [1...N], Source\_url)

-- A Non primary key attribute has a multi-valued attribute, failed 1NF

- Vaccine\_list (Vaccine\_ID, Vaccine\_Name) -> 3NF

### 3. Decomposing into 3NF relations

Using Text to Columns option in Microsoft Excel, I have separated multi-valued attributes into columns.

After splitting the comma separated values in Vaccine column, I used Pivot-Longer function to change to columns into row, so all the individual vaccines used will be in each tuple for the respective location and date

- Location\_list (ISO\_Code, Locations) -> 3NF
- Locations (ISO\_Code\*, LastObservationDate, Vaccines, Source\_Name, Source\_Website) -> 3NF
- Vaccinations (ISO\_Code\*, Date, TotalVaccinations, Total\_Vaccinations\_per\_hundred, People\_vaccinated, People\_vaccinated\_per\_hundred, People\_fully\_vaccinated, People\_fully\_vaccinated\_per\_hundred, Total\_Boosters, Total\_Boosters\_per\_hundred, Daily\_Vaccinations\_raw, Daily\_Vaccinations, Daily\_Vaccinations\_per\_millions, Daily\_People\_Vaccinated, Daily\_People\_Vaccinated\_per\_hundred) -> 3NF
- Vaccinations\_by\_manufacturers (ISO\_Code\*, Date, Vaccine, Total\_Vaccinations) -> 3NF
- Vaccinations\_by\_age\_group (ISO\_Code\*, Date, Age\_group, People\_Vaccinated\_per\_hundred, People\_Fully\_Vaccinated\_per\_hundred, People\_with\_booster\_per\_hundred) -> 3NF
- US\_State\_Vaccinations (ISO\_Code\*, 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) -> 3NF
- Country's\_Daily\_Observation (ISO\_Code\*, Date, Vaccine, Source\_url)
- Vaccine\_list (Vaccine\_ID, Vaccine\_Name) -> 3NF

#### 4. Final Relationship Schema:

I have created a relation between the entities which has the vaccine names with the Vaccine\_List entity and replaced all the vaccine names with Vaccine\_ID and link the relation

Below given is the Final Relationship Schema:

*Location\_list (ISO\_Code, Locations)*

*Vaccine\_list (Vaccine\_ID, Vaccine\_Name)*

*Locations (ISO\_Code\*, LastObservationDate, Vaccine\_ID, Source\_Name, Source\_Website)*

*Vaccinations (ISO\_Code\*, Date, TotalVaccinations, Total\_Vaccinations\_per\_hundred, People\_vaccinated, People\_vaccinated\_per\_hundred, People\_fully\_vaccinated, People\_fully\_vaccinated\_per\_hundred, Total\_Boosters, Total\_Boosters\_per\_hundred, Daily\_Vaccinations\_raw, Daily\_Vaccinations, Daily\_Vaccinations\_per\_millions, Daily\_People\_Vaccinated, Daily\_People\_Vaccinated\_per\_hundred)*

*Vaccinations\_by\_manufacturers (ISO\_Code\*, Date, Vaccine\_ID, Total\_Vaccinations)*

*Vaccinations\_by\_age\_group (ISO\_Code\*, Date, Age\_group, People\_Vaccinated\_per\_hundred, People\_Fully\_Vaccinated\_per\_hundred, People\_with\_booster\_per\_hundred)*

*US\_State\_Vaccinations (ISO\_Code\*, 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)*

*Country's\_Daily\_Observation (ISO\_Code\*, Date, Vaccine\_ID, Source\_url)*