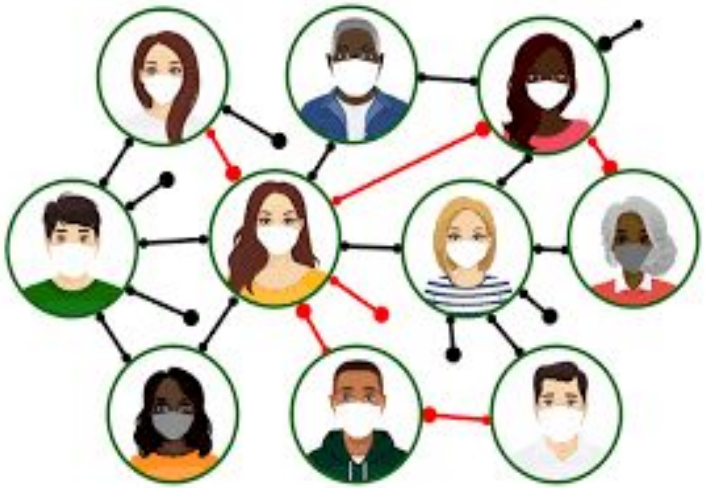
Database Design for COVID-19 Contact Tracing



Project Group 6

Mohini Patil - 015359188
Shwetarani - 015277964
Samhitha Upadhyaya - 015276092

Introduction

- Contact Tracing, a core disease control measure employed by respective health department, which is a key strategy for preventing further spread of Covid-19.
- Contact tracing for Covid-19 requires to identify people who have been in contact with the infected person and monitor them.
- The aim is to apply the concepts of database design and modeling to create a Contact Tracing database for Covid-19 in order to store different information related to users using contact tracing application.

Overview of Project

- The main idea of our project is to keep track of information about people who have come in contact with the person infected by Covid-19.
- The GPS location tracker is used to track the user’s location. Additionally, information about labs and tests conducted in particular cities is maintained.
- Users are tracked to check if they have come in contact with an infected user based on tests conducted in a lab at a particular city.

Overview of Project , cont’d

- Health track admin is the one who can see the information about users and they get notified if the user is infected, in a region where the health check admin holds responsibility.

The risk of having contracted Covid-19 depends on the following factors:

- Individuals having medical conditions which compromises their immunity.
- Individual’s travel history within a given period of time.
- Individuals coming in contact with other infected person based on GPS tracking.

Objectives

- Based on infection status and immunity compromised of users, sending an alert message to them.
- Based on different Tests conducted in the lab, we want to find a list of labs that provide more than one immunity test.
- Based on GPS coordinates of the users, we can identify the locations of the infected persons.
- Based on the travel history and the infection status, we want to determine the country travelled by an infected person.
- To find out non-infected users and their living situation who have come in contact with an infected person (CONTACT TRACING) based on GPS location.
- Using infection status and address of the user we calculate the number of infected cases per city.

Database Initial Study

The End users of COVID-19 Contact Tracer application are users who subscribe for the application. They are regularly tracked if they have come in contact with infected persons and Health officials who analyse the data, and will notify people about whether they are at high risk of having contracted COVID-19 disease.

- There should be an App user who is categorized as User and Health Track Admin
- The User who is general public should be able to access the contact tracing application to fill in necessary information & get notification about his/her infection status.
- The Health Track Admin should be able to access user’s data and monitor infected people within 30km radius of their location coordinates.
- There should be GPS coordinates which will give the user’s location with an accuracy of 100 meters.
- There should be a lab which stores the information about user’s covid-19 Test Results.

Database Design

- The Top Down Design Approach is followed, where ER Model is designed by defining Entities and their attributes, along with relationships between different entities and Relational Schema is designed.
- Defined Business rules
- Developed Conceptual & Logical Model
- Defined Relational Schema
- Developed Physical Model

Entity Relationship Table

ENTITY	RELATIONSHIP	CONNECTIVITY	ENTITY
ADDRESS	lives at	1:M	USER
USER	takes	1:M	TEST
LAB	conducts	1:M	TEST
USER	tracks	M:N	GPS_LOCATION
USER	has	1:M	TRAVEL_HISTORY
APP_USER	receives	1:M	NOTIFICATION
APP_USER	is a	1:1	USER
APP_USER	is a	1:1	HEALTH_TRACK_ADMIN
ADDRESS	loacted at	1:M	LAB

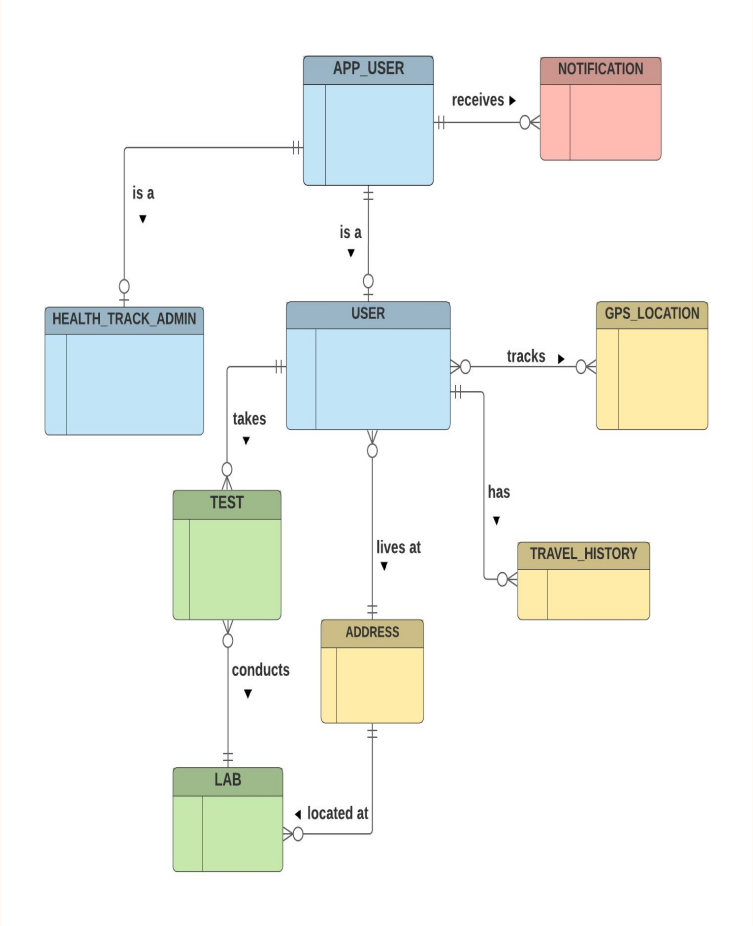
Business Rules

- An APP_USER can be a HEALTH_TRACK_ADMIN and each HEALTH_TRACK_ADMIN is an app user.
- An APP_USER can be a USER and each USER belongs to an app user.
- Each USER can have one address and each ADDRESS belongs to many users.
- Each USER can take many tests and each TEST is carried out on a particular user.
- A Lab can conduct many tests and each TEST is carried out in one lab.
- Each USER can have many tracking GPS locations and each GPS location is tracked by many users.
- Each USER can have many travel histories and each TRAVEL HISTORY belongs to one user.
- An APP_USER can receive many notifications about his risk of contracting COVID-19 and each NOTIFICATION is received by one app_user.
- Each LAB is located at a particular address and each ADDRESS belongs to many labs.

Conceptual Model

The Conceptual Design is a high-level design process which involves describing entities, relationships for COVID-19 Contact Tracer Database.

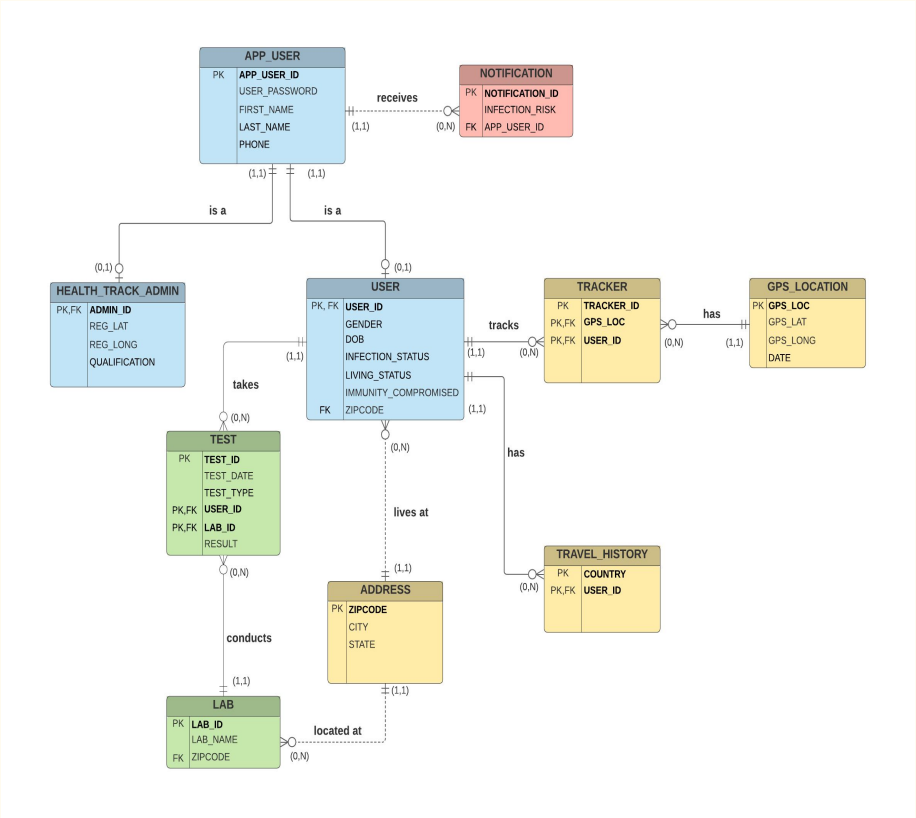
- Defining all the entities.
- Defining relationships between entities.



Logical Model

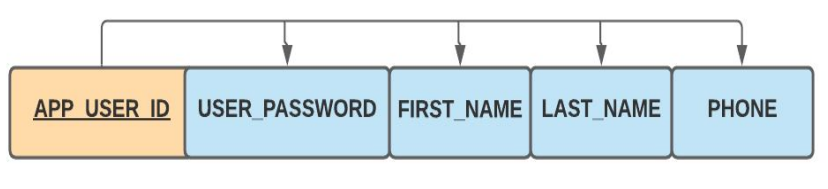
The **logical** data model adds further information to the **conceptual** data model elements.

- Defined all attributes for the entities, identified Primary keys & Foreign keys.
- Defining all the relationships, cardinalities between entities.
- Converting M:N relationships to 1:M
- Identified strong and weak relationships.



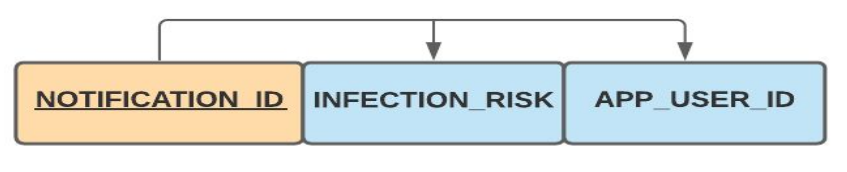
Proof using functional dependencies to show that our tables are in 3NF(3rd NORMAL FORM)

APP_USER



APP_USER_ID --->{USER_PASSWORD , FIRST_NAME, LAST_NAME, PHONE}

NOTIFICATION



NOTIFICATION_ID ---> {INFECTION_RISK , APP_USER_ID}

Relational Schema

APP_USER(**APP_USER_ID**, USER_PASSWORD, FIRST_NAME, LAST_NAME, PHONE)

HEALTH_TRACK_ADMIN (**ADMIN_ID**, REG_LAT, REG_LONG, QUALIFICATION)

USER (**USER_ID**, GENDER, DOB, INFECTION_STATUS,LIVING_STATUS,IMMUNITY_COMPROMISED, ZIPCODE)

GPS_LOCATION(**GPS_LOC**,GPS_LAT,GPS_LONG,DATE)

ADDRESS (**ZIPCODE**, CITY, STATE)

LAB (**LAB_ID**, LAB_NAME, ZIPCODE)

TEST (**TEST_ID**, TEST_TYPE, TEST_DATE,**USER_ID**, **LAB_ID**, RESULT)

NOTIFICATION(**NOTIFICATION_ID**,INFECTION_RISK,APP_USER_ID)

TRACKER(**TRACKER_ID**, **GPS_LOC,USER_ID**)

TRAVEL_HISTORY (**COUNTRY**, **USER_ID**)

Constraints

The translation of the conceptual model into a logical model also requires a definition of the attribute domains and appropriate constraints.

Primary key should be unique and not null

User should have Unique Phone Number

Infection Status:{infected,suspected,not infected}

Living Status:{group,isolated}

Infection Risk:{high,low}

Test Type:{Antibody,RT-PCR}

Physical Design

A physical data model represents relational data objects and used to generate DDL statements.

- Defined table names, column names, column constraints
- Defined attribute data types, primary key, foreign key for table

Security and Access control:

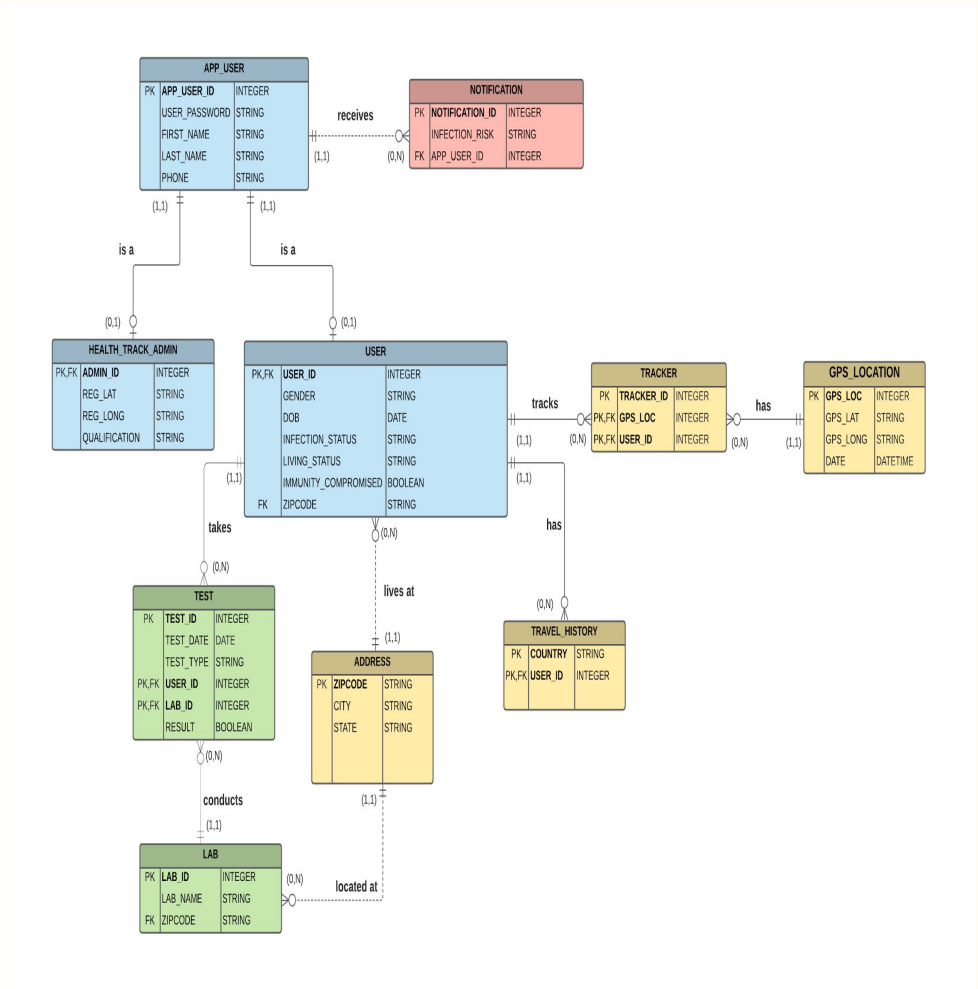
Database owner can perform all the configuration and maintenance activities on the database.

Database admin can have CREATE , READ , UPDATE and DELETE access.

Data analysts can only have access to read the table's data which means, he has granted access to SELECT on Database.

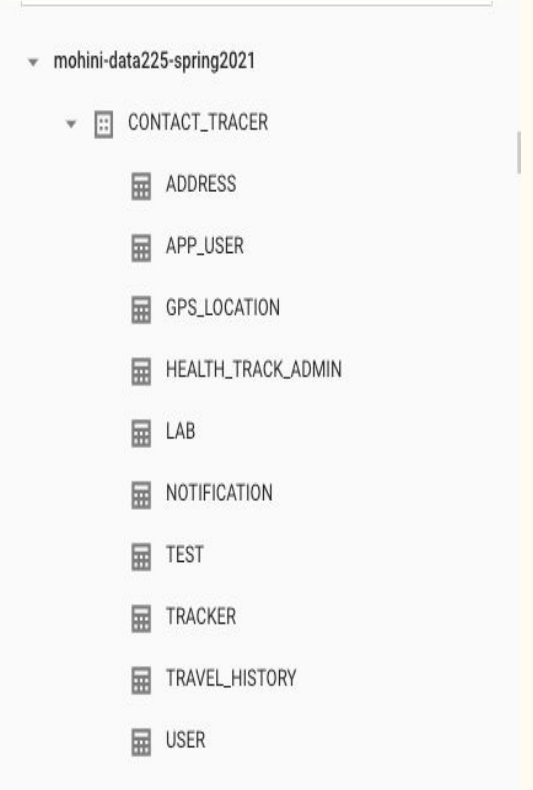
Application user can have only write access to the database, they can only enter the data.

Health_Track_Admin can have access to user's information



DBMS Selection - BigQuery

- BigQuery works great with all sizes of data up to several Petabytes.
- As our database design is for contact tracing mobile applications, if the popularity of the application grows, BigQuery can handle large amounts of data easily and also we can analyze data in BigQuery.
- By using Physical design, we created all the tables and loaded data by using generatedata.com



Testing & Evaluation

Testing and Evaluation using CRUD operations.

C: CREATE - In this operation, we usually test for creating or inserting values into a table.

R: RETRIEVE - This operation is for finding or retrieving information from the database. (i.e SELECT)

U: UPDATE - UPDATE operation takes place, While doing Edit or Modifying an existing record.

D: DELETE - This operation is used to delete the existing record.

Queries (Query Statements defined for Objectives)

- Using SELECT CASE/WHEN, based on infection status and immunity compromised of users, sending an alert message to them.

```
SELECT USER_ID, INFECTION_STATUS, IMMUNITY_COMPROMISED,

CASE WHEN INFECTION_STATUS = 'suspected' AND IMMUNITY_COMPROMISED = TRUE

THEN 'This user is at high risk'

WHEN INFECTION_STATUS = 'suspected' AND IMMUNITY_COMPROMISED = FALSE

THEN 'This user is at moderate risk'

WHEN INFECTION_STATUS = 'infected'

THEN 'This user is Covid-19 positive'

WHEN INFECTION_STATUS = 'not-infected' AND IMMUNITY_COMPROMISED = TRUE

THEN 'This user is at low risk'

ELSE 'This User is safe'

END AS Alert

FROM CONTACT_TRACER.USER
```

Query complete (0.0 sec elapsed, cached)

Job information | Results | JSON | Execution details

Row	USER_ID	INFECTION_STATUS	IMMUNITY_COMPROMISED	Alert
1	51	infected	false	This user is Covid-19 positive
2	45	infected	false	This user is Covid-19 positive
3	16	infected	false	This user is Covid-19 positive
4	42	infected	false	This user is Covid-19 positive
5	20	infected	false	This user is Covid-19 positive
6	52	suspected	false	This user is at moderate risk
7	49	suspected	false	This user is at moderate risk
8	27	suspected	false	This user is at moderate risk
9	12	not-infected	false	This User is safe
10	30	not-infected	false	This User is safe
11	18	not-infected	false	This User is safe
12	17	not-infected	false	This User is safe
13	39	not-infected	false	This User is safe
14	28	not-infected	false	This User is safe
15	38	not-infected	false	This User is safe
16	8	not-infected	false	This User is safe
17	68	infected	false	This user is Covid-19 positive
18	122	infected	false	This user is Covid-19 positive
19	80	infected	false	This user is Covid-19 positive
20	142	infected	false	This user is Covid-19 positive
21	119	infected	false	This user is Covid-19 positive
22	94	infected	false	This user is Covid-19 positive

Queries, cont'd

- Identify the GPS coordinates of the infected person

```
SELECT U.USER_ID, G.GPS_LAT, G.GPS_LONG

FROM `mohini-data225-spring2021.CONTACT_TRACER.GPS_LOCATION` AS G

INNER JOIN `mohini-data225-spring2021.CONTACT_TRACER.TRACKER` AS T

ON G.GPS_LOC = T.GPS_LOC

INNER JOIN `mohini-data225-spring2021.CONTACT_TRACER.USER` AS U

ON T.USER_ID = U.USER_ID

WHERE U.INFECTION_STATUS = "infected"
```

Query complete (0.4 sec elapsed, 27.7 KB processed)

Job information | Results | JSON | Execution det

Row	USER_ID	GPS_LAT	GPS_LONG
1	32	45.37975	-160.22205
2	47	28.94	-18.613
3	20	72.42477	-117.69821
4	50	-84.01231	-176.84174
5	15	60.19581	-149.34752
6	32	-55.14301	1.18273
7	47	21.26325	-88.96061
8	11	25.34848	-154.75583
9	47	-16.785	29.09286
10	42	-73.57437	-14.8258
11	47	-69.90588	-92.48282
12	50	-69.90588	-92.48282
13	11	-84.47124	-32.95661
14	1	47.43498	-19.29981
15	15	47.43498	-19.29981
16	42	24.75319	59.71643
17	43	-25.04256	-144.8108

Queries, cont'd

- Number of infected cases per city

```
SELECT COUNT(U.USER_ID) AS infected_case_count, A.CITY

FROM CONTACT_TRACER.USER AS U

INNER JOIN CONTACT_TRACER.ADDRESS AS A

ON U.ZIPCODE = A.ZIPCODE

WHERE U.INFECTION_STATUS = 'infected'

GROUP BY A.CITY
```

Query complete (0.2 sec elapsed, cached)

Job information | Results | JSON | Execution

Row	infected_case_count	CITY
1	1	Kapolei
2	1	San Antonio
3	1	Grand Island
4	2	Helena
5	10	Flint
6	1	South Portland
7	1	Savannah
8	1	Ketchikan
9	1	Norfolk
10	1	New Orleans

View

- Created a view to display the user's test information like which test user took from which lab and the result of that test.

```
CREATE VIEW CONTACT_TRACER.USERDATA AS(SELECT USER.USER_ID AS UserID,APP_USER.FIRST_NAME as
UserName,LAB.LAB_NAME AS LabName, TEST.TEST_TYPE AS TestType , TEST.TEST_DATE as TestDate,
ADDRESS.STATE as State,

CASE WHEN TEST.RESULT = False THEN 'Negative'

ELSE 'Positive'

END AS TestResult

FROM CONTACT_TRACER.USER , CONTACT_TRACER.TEST ,
CONTACT_TRACER.LAB,CONTACT_TRACER.ADDRESS,CONTACT_TRACER.APP_USER

WHERE LAB.LAB_ID= TEST.LAB_ID AND

TEST.USER_ID = USER.USER_ID AND

USER.ZIPCODE=ADDRESS.ZIPCODE AND

USER.USER_ID=APP_USER.APP_USER_ID

ORDER BY UserID);
```

Query complete (0.4 sec elapsed, 39.2 KB processed)

Job information | Results | JSON | Execution details

Row	UserID	UserName	LabName	TestType	TestDate	State	TestResult
1	1	Lucius	sapien.	Antibody	2020-05-13	Michigan	Negative
2	8	Upton	dignissim	RT-PCR	2019-10-18	Virginia	Negative
3	10	Ronan	rhoncus.	RT-PCR	2020-01-23	Arkansas	Positive
4	11	Clayton	libero.	Antibody	2019-12-17	Michigan	Negative
5	19	Fulton	Utrogger	RT-PCR	2019-12-09	NE	Negative
6	20	Otto	utstar	RT-PCR	2019-06-06	Michigan	Negative
7	28	Rafael	aliquet	RT-PCR	2020-01-09	MT	Positive
8	29	Fitzgerald	Fusce	RT-PCR	2019-05-21	MT	Negative
9	33	Xavier	augue	RT-PCR	2019-06-17	FL	Negative
10	46	Octavius	Fusce	RT-PCR	2019-09-27	Texas	Negative

Stored Procedure

- Created a stored procedure to get the number of infected cases for particular state.

```
CREATE OR REPLACE PROCEDURE CONTACT_TRACER.GetCovidCases (IN STATE_IN STRING, OUT COUNT INT64)
BEGIN
SET COUNT = (SELECT COUNT(U.USER_ID)
FROM CONTACT_TRACER.USER AS U
INNER JOIN CONTACT_TRACER.ADDRESS AS A
ON U.ZIPCODE = A.ZIPCODE
WHERE U.INFECTION_STATUS = 'infected' AND
A.STATE = STATE_IN
);
END;
DECLARE COUNT INT64;
DECLARE STATE_NAME STRING DEFAULT 'Michigan';
CALL CONTACT_TRACER.GetCovidCases(STATE_NAME, COUNT);
SELECT STATE_NAME, COUNT
```

Job information | Results | JSON

Row	STATE_NAME	COUNT
1	Michigan	10

Conclusion

- The proposed system will help in controlling the spread of COVID-19 disease when implemented as a complete system i.e COVID-19 contact Tracing .
- To create the Contact Tracing Database, first we analyzed the situation, we defined business rules and designed models, then loaded the data into the database and then tested and evaluated.
- The proposal states that by using simple applications one can create a public health monitoring system. In conclusion, we believe that implementing such a database system will automate the process of Contact Tracing.

Future Work

- The database designed for COVID-19 Contact tracing can be used for generic contact tracing and for other analytical use cases as well.
- In addition, people surveys can be conducted for users where they are asked to update about their symptoms, body temperature which can then be used as added information to track people infected with disease.

THANK YOU