Tracking and Managing Risk Exposure using Test Results and Location Based Services

Fine tuning operational readiness for Citizens as well as Government and Enforcement Agencies

Introduction

Contagious diseases caused by novel viruses are fast spreading, have high fatality rates and have a big impact on society and economy as witnessed by the current unprecedented situations caused by COVID-19 around the world.

While the current brute force methods used to contain the spread is helping, it is still causing panic, stress and increased anxiety among the common public and started to alter the social behavior of people. Pandemic like this is now de-stabilising the socio-economic status of countries leading to increased rate of close of business and unemployment rates. The problem is magnified by the fact that people can be the carriers way before the actual symptoms show up and in some cases people who are carriers are asymptomatic and exhibit no symptoms of the disease.

Based on the actions of past few months across different countries, it is clear that there is a big gap in terms of insights into meaningful and actionable data available for the government officials, law enforcement authorities and healthcare providers, to make informed decisions and better manage the operational readiness in a given geography.

We must respond not just as a nation's fighting an enemy but as a species fighting for survival. There are two critical steps in fighting this war:

1. Create awareness about the exposure risk and provide necessary tools for people to self Quarantine when getting exposed.
2. Build more accurate datasets and systems to predict the location where the infection rate could increase more and plan the distribution of short supplied medical and other critical resources on priority.

Our proposed solution addresses these problems by connecting the user’s location tracking info and COVID Test result data to build necessary insights and tools to help tackle the situation.

Use Cases

Following are the use cases that this solution addresses:

U#	Description	Comments
U1	Provide meaningful and actionable insights to improve operational readiness in a given geography.	Provide meaningful near real-time data based projections that shows the number of people who are at risk (exposed to the disease) in a given geography (for example: Chicago) and the timeline by when these “at risk” people will need medical help. This information would aid authorities (Governors and Planners) to make data driven decisions on the following: 1) Capacity planning (hospital beds, medical equipment, staff, …) 2) Prevent further spread (focus quarantine efforts for at risk people to stop further spread, like entire city/county shutdown order) **Beneficiary: Government, Law enforcement and HealthCare providers**
U2	Determine if a given commercial location is safe to visit at a given point in time.	Provide meaningful near real-time insights to regular people who are shopping for essentials (for example: grocery store) as to whether the shop they plan to enter is safe based on the following: 1) Are “at risk” people currently in the shop 2) Was the shop visited by a person who tested positive and when and predict whether the establishment is currently contagious 3) When was the location last sanitized Provide insights into shop owners/management for things like when to sanitize outside normal sanitization schedule. **Beneficiary: Citizens, Business Establishments & Supply Chain**
U3	Notify people if they are at increased risk	Provide near real-time notification to people who are at increased risk due to a possible exposure they are not aware of, so that they can quarantine themselves in a timely manner and not put their near and dear ones at risk, and take precautionary measures as advised by CDC. **Beneficiary: Citizens**

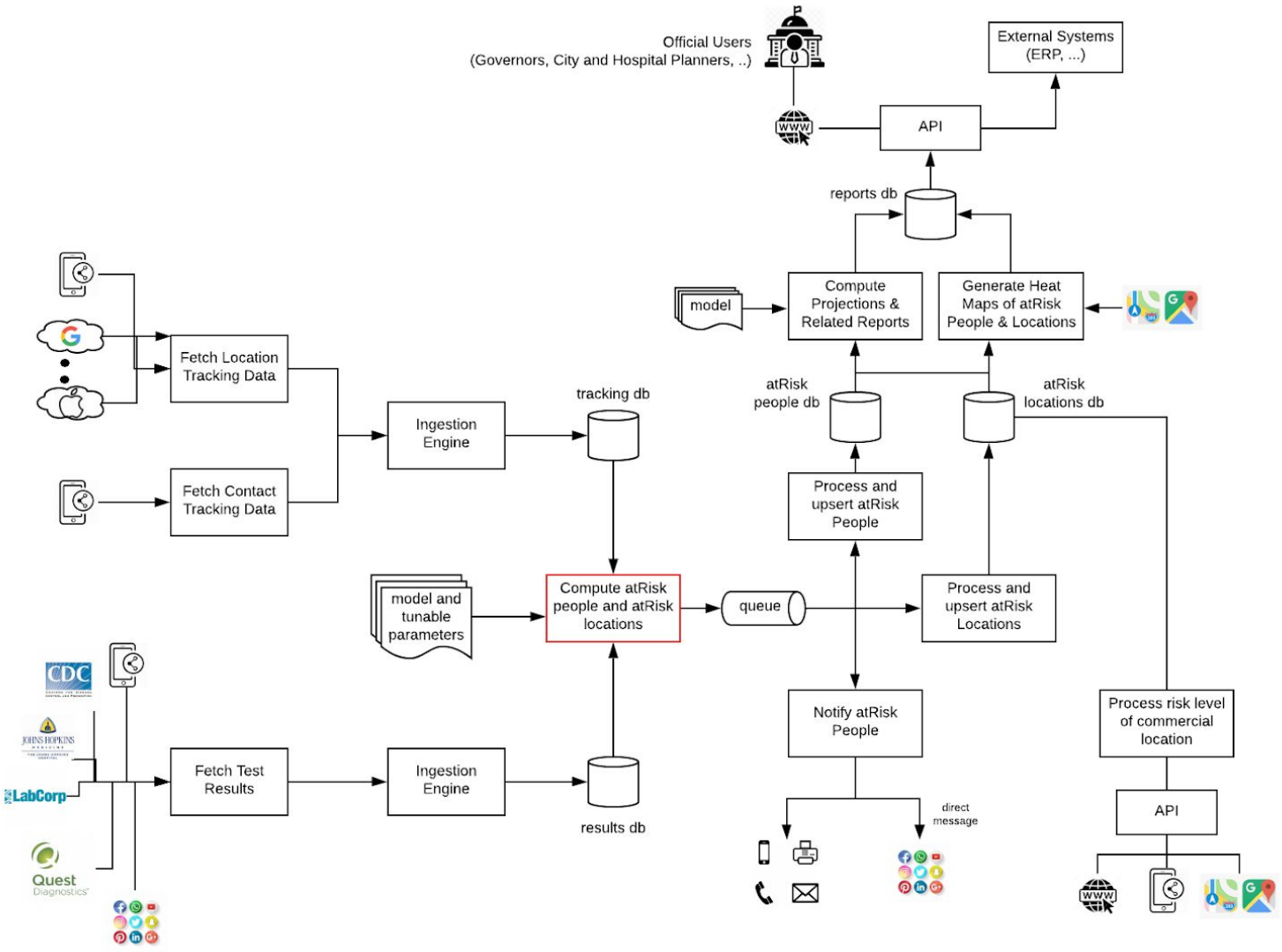
Dependencies

1. People have to voluntarily agree to provide location tracking services to be enabled on their smartphones or other location tracking devices. This kind of data sharing can have serious privacy concerns. The solution proposed will need to have safeguards to ensure that data isn't abused by actors with nefarious intents.
2. HealthCare authorities, such as CDC and Hospitals sharing/feeding the test results to the system for real time tracking
3. Given the times are desperate, the government needs to take equally desperate measures and pass a regulation to allow collection and usage of location tracking data from everyone in a given geography. Such a bill will also require provisions that limit its powers to a short-term time frame till the crisis is over. It will also need strict regulations that will ensure data collection is stopped as soon as the crisis is over and also regulations to ensure that the data is purged once its job is done.
4. Commercial locations that are providing essential services have to agree to provide data on sanitization schedules.

Notes:

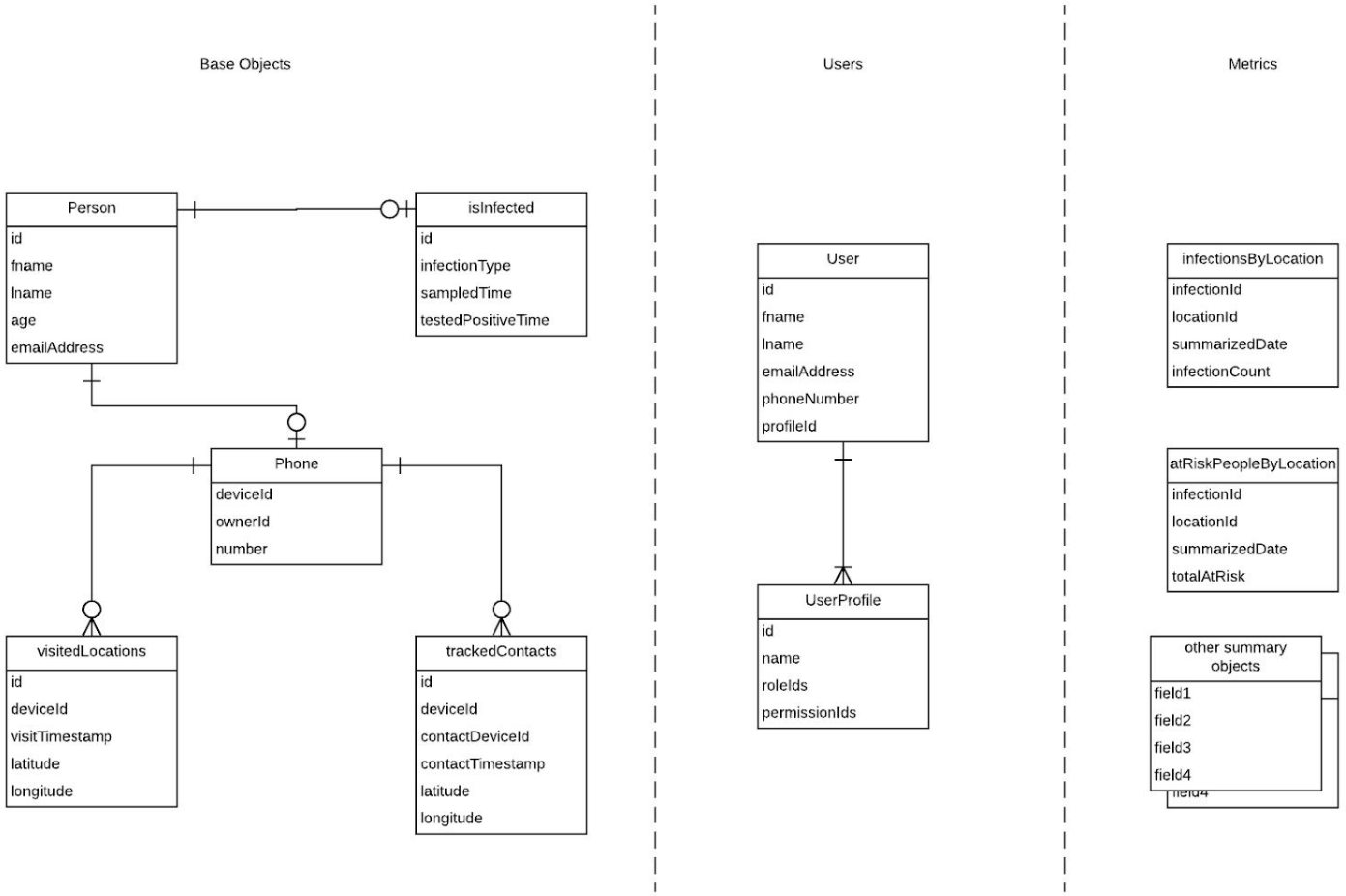
1. The system will keep the location tracking, contact tracking, and test result datasets for the minimum period of time required to generate the required insights. Based on incubation period this is expected to be about four weeks for Covid19 infections.
2. There are privacy concerns around the datasets that we need to collect for this solution to work. As mentioned in the dependencies above, if these concerns are not resolved in a given geography then this solution won't work.
3. This solution will not expose personally identifiable data to non-owners at any point in time.
4. Heatmaps, and metrics around at risk people in a given geography are available only for people with appropriate access level. This system assumes no responsibility in how the data is used by the authorities and what actions they take based on this data.
5. Data is encrypted at rest and in-flight.

Architecture



Data Model

Following are the basic objects in the system for MVP 0.1/PoC:



***Note**: There are several other objects that will be needed as part of the actual MVP/PoC and will be addressed in the detailed design phase.*

Design

Following are some of the critial parameters that influences the computation of atRisk people and locations:

#	name	type	description
1	proximity	distance in meters	physical distance within which the risk of exposure is elevated
2	depth	number	propagation depth that should be computed while determining people at risk
3	lastSanitized	timestamp	UTC timestamp when a given location was last sanitized
4	incubationTime	duration in hours	time from exposure to the time the exposed person is considered to be contagious
5	carrierAfterRecoveryTime	duration in hours	duration of time a fully recovered person is still considered to be contagious and carrier of the disease
6	carrierTime	duration in hours	duration of time before the sampleTime for which the person is considered to be a carrier and contagious
7	bufferTime	duration in hours	additional time buffer to be used to determine since when a person was carrier of the disease
8	pathogenAliveInEnvTime	duration in hours	duration of time the pathogen could be alive in the environment (surfaces, air, water, …) outside the host body
9	sampleTime	timestamp	UTC timestamp when sample was taken for testing
10	resultTime	timestamp	UTC timestamp when the test result was concluded
11	recoveredTime	timestamp	UTC timestamp when the person is declared to be recovered from the disease (based on retest result)
12	carrierEpochTime	timestamp	UTC timestamp since when a given person is considered to be contagious and carrier of the disease carrierEpochTime = sampleTime - carrierTime - bufferTime
13	isInfected	boolean	true=infected, false=notInfected -- based on test result
14	isCarrier	boolean	true=infected, false=notInfected -- based on test result

Data Flows

The project assumes that these datasets are made available from relevant partners and can be processed and stored on AWS without any legal or other restrictions, whatsoever. While this is not an issue for PoC/MVP0.1, we still need to get relevant approvals from relevant authorities and pair up with partners to be able to take this forward.

1) **Location Tracking Data**: Location tracking data is fetched using an app that is tracking the locations visited over time, or by partnering with providers who already have this data (google, apple, facebook, …). This data is fetched in bulk periodically, is processed through the ingestion engine and stored in various internal formats in the trackingDB store. Data storage is optimized for read-heavy workloads.
2) **Contact Tracking Data**: Contact tracking is done by peer-to-peer methodology using an app running on a smart device. The app exchanges contact signature and timestamp over bluetooth every time it detects another device within its proximity range. This data is stored in the smart device and exported to the cloud periodically from where it is then fetched in bulk to our system. This data is processed through the ingestion engine and stored in various internal formats in the trackingDB store. Data storage is optimized for read-heavy workloads.
3) **Test Results Data**: Test results data is fetched from authoritative entities that keeps track of individual test results. This data is processed by the ingestion engine and stored in various internal formats in the resultsDB store that is optimized for read-heavy workloads.
4) **AtRiskPeople & AtRiskLocation Computation**: The data from trackingDB and resultsDB is consumed by the algorithm that computes the list of atRiskPeople (see algorithm details in next section) and also computes atRiskLocations based on movements of atRiskPeople. This information is computed periodically and is deduplicated and written to an internal bus.
5) **Notification Service**: This service consumes the atRiskPeople data from the internal queue, retrieves contact information of these individuals and sends out notifications on one or more preferred channels using specified templates. These templates include links to recommended guidelines and care instructions on what to do next.
6) **LocationStatusAPI**: APIs are made available to check the risk status of a given commercial location at a given point in time. Basic intent of this API is for people/applications to be able to find if a given commercial location is safe to enter at a given point in time. The implementation checks to see if atRisk people are currently present or have recently visited this location to determine the risk status of this location. Additionally the contact tracking status application can be updated with riskStatus of a given individual and be used to flag a warning if proximity with atRisk person is detected, however, this approach is not advisable as it pinpoints atRisk people and this could put them in danger in some societies.
7) **Report Jobs**: Various summary metrics, heat maps, and projections are generated by periodic jobs that consume atRiskPeople and atRiskLocations data and stores these reports in reportsDB.
8) **ReportsAPI**: APIs are made available for authorized users to obtain various reports and metrics from this system. These APIs are limited to official users (Government, Law Enforcement, Hospital Planners, …) and not to regular end-users.
9) **Alerts**: Additional alerting system could be setup to warn official users when metrics breach certain configured levels and these alerts flow through the notification service but on priority queues.

Computing AtRiskPeople and AtRiskLocations:
The design assumes that location and contact tracking data and test results data is constantly (every few minutes) fetched, processed and stored in the system.

AtRiskPeople are computed in the following two ways:

1) By tracking movements of people whose test results just came back as positive (isInfected=true)
2) By tracking movements of people who are already categorized as atRiskPeople.

For people whose test results just came back as positive:

1. Initialize masterListOfAtRiskPeople
2. Determine the carrierEpochTime
3. Get all the locations visited (along with visit timestamp) by this person since carrierEpochTime
4. For each location
    a. Get all the other people who happened to be in the proximity at the same time (first circle)
    b. Dedup this list (if the same contact appears multiple times and multiple locations, keep the oldest exposure time) w.r.t. masterListOfAtRiskPeople and append to the masterList.
5. For each person in the list computed from step-3:
    a. Set carrierEpochTime = exposureTime + incubationTime
    b. Repeat steps-2 and 3 above to get subsequent circles of people who were exposed to the pathogen
6. Iteratively repeat step-5 till the list is exhausted or till carrierEpochTime after applying timeDecay becomes greater than or equal to current time.

In addition to location tracking, we prepare another master list using a similar approach with contact tracking data and then dedup these lists to get the final list.

Similar approach is taken based on movements of people who are already categorized as atRisk.

AtRiskLocations are determined based on current presence or recent visit (within pathogenAliveInEnvTime) of atRiskPeople at that specified location.

Future Enhancements

- Geofencing and Alerting for emergency situations
    - Crime in progress,
    - Environmental Pollution
    - Pupil tracking within school zones
- ML based projections
    - Use additional data points (age, gender, pre-existing conditions, heuristics, demographics, …) as signals to machine learning algorithms for determining various projections of sick people over time in a given geography

Team

#	Name	Email
1	Palani Vel	topalanivel@gmail.com
2	Kathiravan Sengodan	skathiravan@gmail.com
3	Vishnu Prasad Ramu	vishnu.mini@gmail.com
4	Pratik Chapadgaonkar	darkrai007prat@gmail.com
5	Kishore Kumar Kurri	dearkishore@gmail.com
6	Vijaya Saradhi Pula	pvsaradhi@gmail.com